FTD-HT-23-659-67

FOREIGN TECHNOLOGY DIVISION

AIR FORCE SYSTEMS COMMAND

CYBERNETICS AND DOCUMENTALISTICS
(COLLECTION OF ARTICLES)

50 YEARS
1917 1967
GOLDEN ANNIVERSARY
FOREIGN TECHNOLOGY DIVISION

21

FTD-HT-23-659-67

UNEDITED ROUGH DRAFT TRANSLATION

CYBERNETICS AND DOCUMENTALISTICS
(COLLECTION OF ARTICLES)

English pages: 196

Translated under: Contract AF33(657)-16408

TT7002135-2140

THIS TRANSLATION IS A RENDITION OF THE ORIGINAL FOREIGN TEXT WITHOUT ANY ANALYTICAL OR EDITORIAL COMMENT. STATEMENTS OR THEORIES ADVOCATED OR IMPLIED ARE THOSE OF THE SOURCE AND DO NOT NECESSARILY REFLECT THE POSITION OR OPINION OF THE FOREIGN TECHNOLOGY DIVISION.	PREPARED BY: TRANSLATION DIVISION FOREIGN TECHNOLOGY DIVISION WP-AFB, OHIO.

1	2	3

This translation was made to provide the users with the basic essentials of the original document in the shortest possible time. It has not been edited to refine or improve the grammatical accuracy, syntax or technical terminology.

Akademiya Nauk SSSR
Nauchnyy Sovet po Kibernetike
Seriya "Organizatsiya i Upravleniye"

KIBERNETIKA
I DOKUMENTALISTIKA

Mekhanizatsiya Protsessa
Nakopleniya, Khraneniya
i Poiska Nauchnoy Informatsii

Izdatel'stvo "Nauka"
Moskva 1966

176 pages

FTD-HT-23-659-67

ITIS INDEX CONTROL FORM

01 Acc Nr	68 Translation Nr	65 X Ref Acc Nr	76 Reel/Frame Nr
TT7002135	FTD-HT-23-659-67		1882 0699

97 Header Clas	63 Clas	64 Control Markings	94 Expansion	40 Ctry Info
UNCL	UNCL, 0	0		UR , US

02 Ctry	03 Ref	04 Yr	05 Vol	06 Iss	07 B. Pg.	45 E. Pg.	10 Date
UR	0000	66	000	000	0005	0034	NONE

Transliterated Title
PROBLEMA DOKUMENTAL'NOY INFORMATSII

09 English Title
THE PROBLEM OF DOCUMENTAL INFORMATION

43 Source AN SSSR. NAUCHNYY SOVET PO KIBERNETIKE. SERIYA "ORGANIZATSIYA I UPRAVLENIYE." KIBERNETIKA I DOKUMENTALISTIKA (RUSSIAN)

42 Author	98 Document Location
VOROB'YEV, G. G.	
16 Co-Author NONE	**47 Subject Codes** 05, 09
16 Co-Author NONE	**39 Topic Tags:** cybernetics, information processing, information storage and retrieval, machine abstracting, computer
16 Co-Author NONE	
16 Co-Author NONE	

ABSTRACT : This article discusses the problems of documental information: the vast amount of information which is being written; the increasing number of authors; the problems of library useage due to roughly cataloged subject indexes, lack of funds, and the importance of scientific articles versus books. The author specifically states the importance of information retrieval and covers this under: documentalistics - a new branch of cybernetics; the theory of documental information systems; mechanization and automation of information systems which covers storage and retrieval of information by the use of a computer; and organization of information services.
English translation: 39 pages.

ITIS INDEX CONTROL FORM

01 Acc Nr	68 Translation Nr	65 X Ref Acc Nr	76 Reel/Frame Nr
TT7002136	FTD-HT-23-659-67		1882 0700

97 Header Clas	63 Clas	64 Control Markings	94 Expansion	40 Ctry Info
UNCL	UNCL, 0	0		UR, US

02 Ctry	03 Ref	04 Yr	05 Vol	06 Iss	07 B. Pg.	45 E. Pg.	10 Date
UR	0000	66	000	000	0035	0080	NONE

Transliterated Title

MIKROFOTOREPRODUKTSIONNYYE METODY SVERTYVANIYA DOKUMENTAL'NOY INFORMATSII

09 English Title

MICROPHOTOREPRODUCTION METHODS OF REDUCING DOCUMENTAL INFORMATION

43 Source AN SSSR. NAUCHNYY SOVET PO KIBERNETIKE. SERIYA "ORGANIZATSIYA I UPRAVLENIYE." KIBERNETIKA I DOKUMENTALISTIKA (RUSSIAN)

42 Author	98 Document Location
LIPKIN, S. A.	
16 Co-Author NONE	**47 Subject Codes** 05, 09, 14
16 Co-Author NONE	**39 Topic Tags:** microphotography, photographic material, data storage, information theory, computer, punched card
16 Co-Author NONE	
16 Co-Author NONE	

ABSTRACT : This article describes microphotoreproduction methods of reducing documental information: 1) reduction of document dimensions; 2) technology of large-scale microphotocopying: photographic materials such as Diazo film and Kalvar photographic film, photochromatic materials, and equipment and technology for mass production of roll microphotocopies and microcards; 3) mechanization and automation of information searching: manual search using negative and positive single- and multi-row microcards; aperture punchcards; systems for highly mechanized automated searches based on the use of microphotocopies of documents; several file searching and filmsorting machines and systems manufactured by U.S. firms; 4) reading and efficient coding of microphotocopies using various types of readers, pocket readers, and reader-copiers manufactured by both U.S. and Soviet firms.
English translation: 47 pages.

ITIS INDEX CONTROL FORM

01 Acc Nr	68 Translation Nr	65 X Ref Acc Nr	76 Reel/Frame Nr
TT7002137	FTD-HT-23-659-67		1882 0701

97 Header Clas	63 Clas	64 Control Markings	94 Expansion	40 Ctry Info
UNCL	UNCL, 0	0		UR

02 Ctry	03 Ref	04 Yr	05 Vol	06 Iss	07 B. Pg.	45 E. Pg.	10 Date
UR	0000	66	000	000	0081	0089	NONE

Transliterated Title PRIMENENIYE KART S KRAYEVOY PERFORATSIYEY DLYA REFERATIVNOY INFORMATSII NA VAZE RZH "METALLURGIYA"

09 English Title USE OF EDGE-PUNCHED CARDS FOR ABSTRACT INFORMATION ON THE BASIS OF THE RZh "METALLURGIYA"

43 Source AN SSSR. NAUCHNYY SOVET PO KIBERNETIKE. SERIYA "ORGANIZATSIYA I UPRAVLENIYE." KIBERNETIKA I DOKUMENTALISTIKA (RUSSIAN)

42 Author	98 Document Location
KOLESNIKOVA, T. P.	

16 Co-Author	47 Subject Codes
SHMELEVA, A. V.	09, 05

16 Co-Author	39 Topic Tags:
NONE	punched card, encoding theory, information storage and retieval, coding

16 Co-Author
NONE

16 Co-Author
NONE

ABSTRACT : This article describes the use of edge-punched cards for abstract information to facilitate the search for current information on metallurgy listed in RZh Metallurgiya. A punchcard model used for this process is shown in the figure below. Both combined and direct coding methods are used in the card catalogs on metallurgy. The combined code is used to encode the bibliographic mutually exclusive lables, and the direct code is used for the parallel subject labels and materials.
English translation: 10 pages.

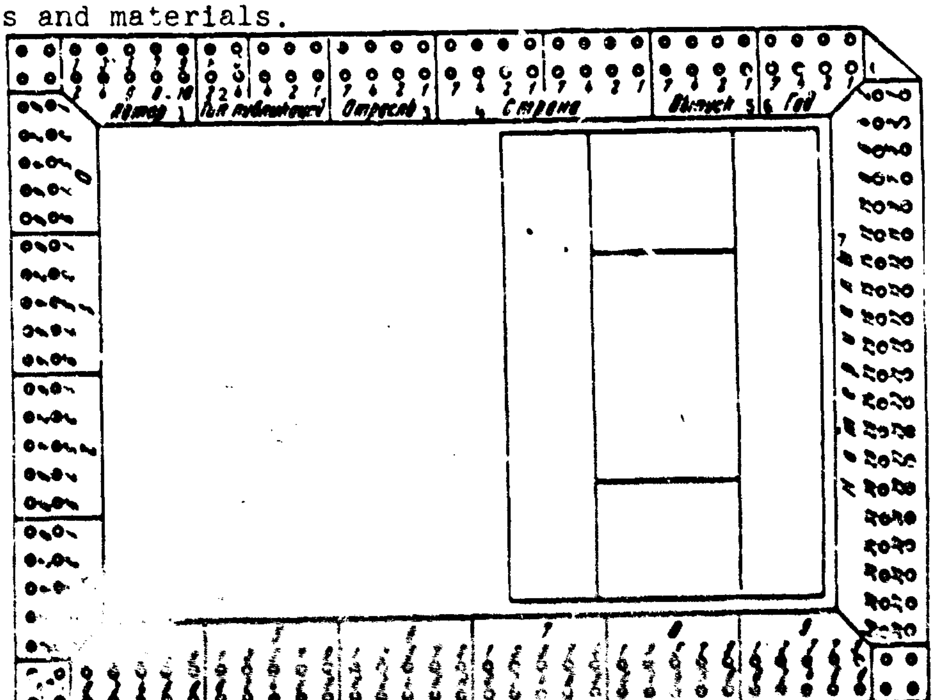


Fig. 1. Punchcard model. 1) Author; 2) publication type; 3) field; 4) country; 5) issue; 6) year; 7) materials.

FTD Form Feb 67 0-90

ITIS INDEX CONTROL FORM

01 Acc Nr	68 Translation Nr	65 X Ref Acc Nr	76 Reel/Frame Nr
TT7002138	FTD-HT-23-659-67		1882 0702

97 Header Clas	63 Clas	64 Control Markings	94 Expansion	40 Ctry Info
UNCL	UNCL, 0	0		UR

02 Ctry	03 Ref	04 Yr	05 Vol	06 Iss	07 B. Pg.	45 E. Pg.	10 Date
UR	0000	66	000	000	0090	0100	NONE

Transliterated Title PUTI PRIMENENIYA INFORMATSIONNO-POISKOVYKH SISTEM V MEDITSINE

09 English Title WAYS OF USING INFORMATION SEARCH SYSTEMS IN MEDICINE

43 Source AN SSSR. NAUCHNYY SOVET PO KIBERNETIKE. SERIYA "ORGANIZATSIYA I UPRAVLENIYE." KIBERNETIKA I DOKUMENTALISTIKA (RUSSIAN)

42 Author	98 Document Location
AMOSOV, N. M.	
16 Co-Author	**47 Subject Codes**
MINTSER, O. P.	09, 05, 06
16 Co-Author	**39 Topic Tags:** information storage and retrieval, medicine, data processing, punched card
NONE	
16 Co-Author	
NONE	
16 Co-Author	
NONE	

ABSTRACT: The following operations may be performed with the use of the information search system described in this article.

1. Search for a clinical precedent. The solution of this problem permits at the very first stages of the system operation the determination of the risk of operative intervention, the modification of a diagnosis, and evaluating the possible severity of the patient condition in the post-operative period.
2. Verify the reliability of several methods of research.
3. Perform statistical and correlation calculations.
4. Determining the efficiency of treatment methods.

Naturally, the suggested scheme is still far from perfect, the widest possible range of doctors and researchers of the scientific and technical information institutes must participate in its development. However, the results of the use of the information search systems in medical practice will not be long in being felt and even at the very first stages they will aid in improving the precision of clinical medicine.
English translation: 13 pages.

ITIS INDEX CONTROL FORM

01 Acc Nr	68 Translation Nr	65 X Ref Acc Nr	76 Reel/Frame Nr
TT7002139	FTD-HT-23-659-67		1882 0703

97 Header Clas	63 Clas	64 Control Markings	94 Expansion	40 Ctry Info
UNCL	UNCL, 0	0		UR

02 Ctry	03 Ref	04 Yr	05 Vol	06 Iss	07 B. Pg.	45 E. Pg.	10 Date
UR	0000	66	000	000	0101	0137	NONE

Transliterated Title RESHENIYE NEKOTORYKH INFORMATSIONNO-LOGICHESKIKH ZADACH OBSHCHEY KHIMII NA ELEKTRONNYKH TSIFROVYKH VYCHISLIT"L'NYKH MASHINAKH

09 English Title SOLUTION OF SOME INFORMATION LOGIC PROBLEMS OF GENERAL CHEMISTRY ON ELECTRONIC DIGITAL COMPUTERS

43 Source AN SSSR. NAUCHNYY SOVET PO KIBERNETIKE. SERIYA "ORGANIZATSIYA I UPRAVLENIYE." KIBERNETIKA I DOKUMENTALISTIKA (RUSSIAN)

42 Author	98 Document Location
SEYFER, A. L.	
16 Co-Author	**47 Subject Codes**
NONE	09, 05, 07
16 Co-Author	**39 Topic Tags:** information storage and retrieval, computer logic, machine translation, information processing, chemical compound
NONE	
16 Co-Author	
NONE	
16 Co-Author	
NONE	

ABSTRACT: Four basic trends in the creation of information logic systems for general chemistry are presented in this article. The automatic information search for chemical compounds on the basis of their physical and chemical properties; the accumulation, processing, and retrieval of information using composition-property diagrams; nomenclature automatic translation; and the information logic methods of calculating properties of compounds. Investigations described in this article were primarily made at the Laboratory of Electrosimulation of the Institute of Scientific Information of the AN SSSR.
English translation: 45 pages.

ITIS INDEX CONTROL FORM

01 Acc Nr	68 Translation Nr	65 X Ref Acc Nr	76 Reel/Frame Nr
TT7002140	FTD-HT-23-659-67		1882 0704

97 Header Clas	63 Clas	64 Control Markings	94 Expansion	40 Ctry Info
UNCL	UNCL, 0	0		UR, US

02 Ctry	03 Ref	04 Yr	05 Vol	06 Iss	07 B. Pg.	45 E. Pg.	10 Date
UR	0000	66	000	000	0138	0149	NONE

Transliterated Title INFORMATSIONNAYA SISTEMA PO MATEMATICHESKOY TEORII EKSPERIMENTA

09 English Title AN INFORMATION SYSTEM ON THE MATHEMATICAL THEORY OF EXPERIMENTATION

43 Source AN SSSR. NAUCHNYY SOVET PO KIBERNETIKE. SERIYA "ORGANIZATSIYA I UPRAVLENIYE" KIBERNETIKA I DOKUMENTALISTIKA (RUSSIAN)

42 Author
NALIMOV, V. V.

98 Document Location

16 Co-Author
ADLER, YU. P.

47 Subject Codes
09, 05, 12

16 Co-Author
GRANOVSKIY, YU. V.

39 Topic Tags: information theory, computer technology, data analysis, statistics, publication

16 Co-Author
ALEKSANDROVA, I. F.

16 Co-Author
VOROB'YEV, G. G.

ABSTRACT: This article presents data for an information system designed for a collective search of new ideas in the field which is termed mathematical theory of experimentation. An important task of this theory will be to find optimal conditions for the operation of technological process. Using the probability approach it is possible to apply a strictly deterministic approach with the subsequent use of variation analysis, dynamic programming, and the Pontryagin maximum principle. Both the probability methods and variational methods of analysis will be used simultaneously for some publications; however for other publications statistical methods and methods of information theory will be utilized. The mathematical theory of experimentation is essentially related to numerical methods of analysis and modern computer technology. The object of this information system is to permit the development of new trends in current publications and also carry out a retrospective search using publications of past years.
English translation: 28 pages.

FTD Form 10 67 0-90

TABLE OF CONTENTS

FROM THE EDITOR

It apparently will not be a great exaggeration to say that the development of science in the near future will to a considerable degree be determined by the organization of the information services. Today we can speak of the appearance of a new discipline – documentalistics, which concerns the questions of the mechanization of the process of collection, storage, and retrieval of information. The development of this discipline may be followed in the individual issues of the abstract journal "Scientific and Engineering Information," published by VINITI of the Academy of Sciences USSR. In 1964 this journal summarized more than 3000 publications relating to documentalistics. Search for scientific information is a complex intellectual process. It is based on several psychological factors: intuition, ability to associate ideas, the sense of satisfaction from the results of successful search, etc. In an attempt to mechanize this process we encounter the "man-machine" problem, one of the basic problems of cybernetics.

The present volume is the first collection of Soviet works on documentalistics. The papers selected for publication have been presented and discussed at seminars on documentalistics of the Scientific Council on Cybernetics of the Academy of Sciences USSR.

THE PROBLEM OF DOCUMENTAL INFORMATION

G.G. Vorob'yev

We live in the century of electronic machines and information processing [1]. The wide usage of the term "information," which we got along without quite easily up until now, apparently is an indication of the appearance of a special problem. Its formulation may be expressed by the words: the human organism can no longer cope with the information being received and transmitted; the creation of external memory organs – libraries and archives – provided for progress only at a definite stage of the development of society; new revolutionary transformations are necessary, whose importance will be no less than that of the appearance of the first books.

Unfortunately, we do not really understand the scale of this problem.

"The accumulation of human experience proceeds with amazing rapidity, but the means which we utilize to struggle through the complex labyrinth to the objective which is of interest to us at a given time remain the same as in the time of sailing vessels" [2, page 7]. "The volume of knowledge accumulated in one form or another appears to some of us in the form of a modern dinosaur. These ancient animals became extinct, and in part their disappearance is ascribed to the circumstance that they were too large: their brain could not control the tremendous body. The volume of information accumulated by man also may become so great that it is impossible to effectively utilize its individual parts" [3]. In many fields the situation is being created when es-

sentially it is easier to discover a new fact or create a new theory than to ascertain that it has not yet been created or derived. "... The problem is not that we are busy with excessive publication as a result of the scope and variety of modern interests, but rather that the volume of the publication is growing far beyond the limits of our true capability of really utilizing these data" [2, page 23].

Books and Journals

Many indices of the development of human society grow exponentially, i.e., they increase ten-fold every 50 years [4]. An example are the

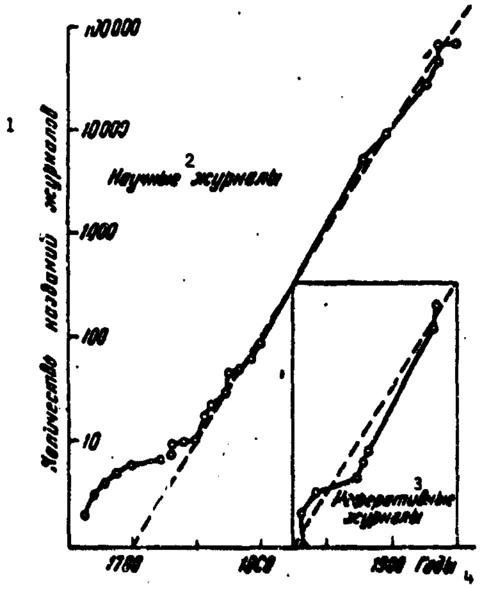


Fig. 1. Exponential growth of number of scientific journals. 1) Number of journal titles; 2) scientific journals; 3) abstract journals; 4) years.

scientific journals, whose number has increased by a factor of 10,000 in 200 years (1750-1950) (Fig. 1).

The annual volume of printed information amounts to 7 billion pages today, including 2-3 million scientific articles (60 million pages), 60,000 book titles, 200,000 descriptions for authors' certificates and

patents, 4 million publications of an economic nature (advertisements, reports, predictions, etc.) [5-8]. Every year 500 new journals on chemistry appear [9] and in the USA alone there are 25 new agricultural journals [10]. During 40 years the All-Union Book Chamber has recorded more than 22 million items of all forms of Soviet printed material [11].

The traditional forms of books and journals do not answer the requirements of today. Let us list the primary deficiencies: the increasing difficulty of publication (excess of the author's supply over the reader's demand), shortening of the volume of scientific articles with the loss of useful information, scattering of the interests of the readers among many journals and the associated decrease of the number of subscribers, discrepancy between the number of readers and the publication volume of articles, publication delay, increase of the gap between the obsolescence of published material and its availability, inconvenience of storage.

In spite of the growth of the number of journals, the number of potential authors is growing even faster. Today's publishing service cannot cope with the stream of manuscripts and attempts to reduce this stream mechanically. For example, the well-known journal "Zavodskaya laboratoriya" [Industrial Laboratory] during 20 years of existence has somewhat increased its size (primarily by use of internal techniques — reduction of margins, smaller type), has cut in half the average article size, and still is forced to reject almost half the manuscripts solely for lack of space [4]. Another journal "Physical Review" during 10 years (1946-1956) increased the size of the annual volume from 1734 to 7347 pages, collects a payment of $25 per page from the authors, and is proposing to limit itself to abstracts of the articles in the future, publishing the latter only on requests of the readers [12].

Thus, we encounter the first contradiction: there are too many journals and too few; the editors in their own way are correct in shortening the material, but at the same time they are doing irreparable damage to scientific and technical information. Let us clarify this last idea by an example. A scientist has conducted in his laboratory a major experiment and has published a lengthy article concerning it, where he has described not only the experimental conditions and the results obtained but also secondary phenomena which he has not been able to explain definitely because of lack of facts. It is obvious that a brief communication will only be of a publicity nature, it can neither be utilized nor verified [13].

Along with the increase of the number of journals there is a reduction of the relative number of subscribers. The reason for this is not so much the cost [14-16], as the specialization and scattering of interests of the readers. On the average, the fields of interest of the user of documental information in 1962 were reduced by half in comparison with the figure for 8 years previously, and will again reduce by a factor of two by 1970 [3]. In each issue of a journal we are interested on the average in two or three articles which we would like to retain, but there are tens or hundreds of such journals [17]. In order to subscribe to all, we would need a lot of money, a lot of available space, and in all cases we would obtain a mass of "waste paper" in which there is invested the labor of the papermakers, printers, binders, and others and which will occupy space, interfering with us in searching for the required information. We might add that the readers of each article differ not only qualitatively, but also quantitatively: hundreds are interested in one article, millions of specialists are interested in another article, but both articles will be located side by side in an issue of the journal and are thus presented to 10,000 subscribers.

We often pride ourselves on modern polygraphic techniques, but we don't stop to think that they are capable of increasing the output but not the speed of publication. The great Russian printer Ivan Fedorov set and printed the text of Evangeliye uchitel'noye in 8 months and 9 days. After 400 years the 11*th* volume of the complete collection of the works of Charles Dickens, not requiring author's correction, required 8 months and 20 days for printing, and the 30*th* volume required more than a year.

But even after publication the article still must follow a long and tortuous path to the person interested in it. This applies particularly to periodic and nonperiodic volumes containing the works of scientific research institutes and the educational institutions where most scientific work is concentrated. Their circulation is small and the abstract services are reluctant to acquire and process them under the pretext that in the course of time they will be republished in the leading journals. This may be why the average accessibility of the literature is extremely low: 75% in physics 10 years after publication, 55% in chemistry, and still less in the other sciences (Fig. 2). We re-

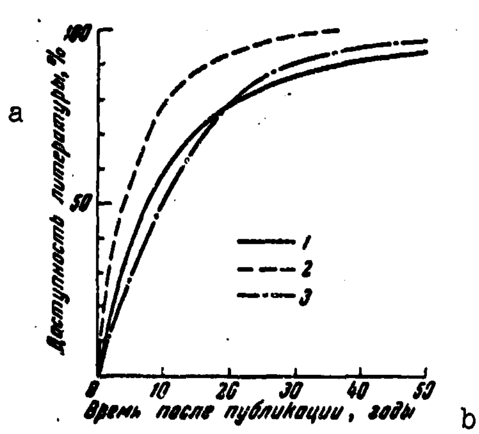


Fig. 2. Accessibility of published scientific and engineering information. 1) Chemistry; 2) physics; 3) mathematics. a) Literature accessibility, %; b) time after publication, years.

call that the USA industry became aware of the development of the Soviet turbine drills only 6-7 years after publication in the Soviet press [16]. But on the average the scientific and engineering literature is out of date after 10-15 years [18].

Abstract Service

In the first half of the twentieth century a second form of information won wide acceptance – the specialized abstract volumes which present a brief summary of the contents of the published scientific articles. However, later on all the deficiencies of the conventional journals were repeated here: delays, unwieldyness, expense, inconvenience in use, small fraction of useful information (for the single subscriber), etc.

The abstract journals quickly filled the shelves of the libraries, becoming a technical problem of their own. One of the largest information services "Chemical Abstracts" from 1907 to 1960 increased the number of chemical journals abstracted annually from 475 to 9800 titles, the number of abstracts to 104,500, and the number of chemical compounds described to 160,000 (including 100,000 new compounds [9]. 14 information services of the USA prepared in 1960 600,000 abstracts and bibliographic descriptions, which amounts to about 10% of the published scientific information [19]. In the USSR in 1953 there was created a combined "Abstract Journal" on the basis of the All-Union Institute of Scientific and Technical Information – VINITI. By 1962 this journal increased the number of articles handled to 781,000 (from 102 countries of the world), still without encompassing everything either with respect to branches of knowledge or with respect to number of publications.

Each year the world spends 200 million dollars on abstract servic-

es [20], but a very small percentage of this sum has any appreciable effect. The reason is that no one has yet decided how to accomplish complete coverage of the literature, whether we must abstract all articles and how many times (in various aspects) to abstract a single article [3]. The fields of interest of the users of documental information are not only narrowing (differentiating) [3], but are also diffusing (integrating) [17] among many scientific and abstract journals and the individual sections of these journals. The article "Statistical processing of the results of isotope measurements of lead in meteorites" simply on the basis of the title must be applicable to the circle of interests of geologists, astronomers, physicists, geophysicists, geochemists, chemists, and mathematicians.

The writing of a universal abstract means nearly complete repetition of the entire article or not writing anything. Even an article which is abstracted in one section of a broad-field journal may be lost for the readers of other sections who do not happen to glance in the neighboring section, do not have enough time for this, or simply do not suspect from the specialized abstract that there is useful information in this article.

Libraries

The situation in the libraries is no less difficult. The total number of accumulated printed works amounts today to 100 million titles, of which nearly 50 million are books and about 10 million patents. The V.I. Lenin State Order of Lenin Library of the USSR [Lenin Library], one of the largest in the world, has 22 million books, journals, and bound volumes of newspapers, which occupy 300 km of bookshelves; and the shelf length increases by 15 km each year [11].

"I believe that the majority of people instinctively do not like

libraries and rate them somewhere after dentists on the basis of the inconveniences which they cause," noted at one time one of the English scientists [22]. The huge number of books and journals, to which, unfortunately, they do not always have direct access, have a depressing effect on the library visitor.

And the library catalogs are lots of trouble and of little use. The author catalog – an intrinsic part of every library – apparently satisfies only the lover of classical fiction, but is nearly completely unnecessary in a scientific library where the reader is interested not in *who* wrote on a given subject, but in *what* has been written on this subject. Another error of the librarian: giving more attention to the books than to articles. As a rule, the books are "in the rear" of science and serve certain definite purposes: as textbooks, popular reviews for expanding the horizon, and as handbooks on the desk of a specialist, but those who are at the "leading edge" need primarily the scientific articles.

The so-called systematic and subject catalogs are compiled using a rough classification scheme on the basis of book titles. To accomplish this there is a two or three-fold duplication of the library card files. But every article and more especially books correspond not to two or three, but to many dozen subjects, and only with a suitable breakdown do they become really effective [17]. According to the calculations of Voigt [30], the users obtain 4-7% of the scientific information with the aid of the library, 35% as a result of independent literature searches, and 6-20% as a result of their own memory, leading them to the required original sources.

The production of card-catalog bibliography in the USSR is accomplished primarily by the All-Union Book Chamber, which produces in a year several hundred thousand cards with a total circulation of several

hundred million copies and which operates 18 catalogs with 9 million cards [11, 23]. The question naturally arises: "Is it advisable to catalog so extensively these materials, can the libraries utilize in their card catalogs 75-80,000 cards per year?" [23]. We can glance in any library and see that the majority of the cards go into the wastepaper basket and in the best of cases (the back side) is used for other purposes.

The primary difficulties in the operation of the libraries lie in the absence of any mechanization [24], the limited funds, area, and frequently in the reluctance to change a system which has been developed over the years. Only these factors can explain the understaffing of many libraries, the orientation of engineers and scientists toward the specialized rather than universal (public) libraries [25], and the low utilization of the materials. For example, more than half the books of the Lenin Library have not been used by anyone, apparently not because these books are not required by anyone, simply the readers do not know of their existence.

Other Documental Information Services

So far we have discussed only the so-called scientific and technical information, whose principal task is the collection, processing, and distribution of two opposing information streams - internal and external, which include information concerning new scientific investigations, disclosures, and inventions. The libraries combine the scientific and technical information with the archive and museum information, where the problem of record keeping is just as important. Today the state archives contain tens and hundreds of millions of documents. If at the present time years are spent in searching for documents and we do not always by any means find what we need if because of lack of

space the older archives have been destroyed prematurely, this means the tasks of the archive operation have not yet been resolved. It is remarkable that with the decrease of the probability of major geographical discoveries there is a continuous increase of the probability of no less important discoveries in our archives [26]. Construction drawings, the most inconvenient form of documentation (folded sheets of large format) occupy a special category of archives. For the construction, for example, of a ship we require from 3 to 5000 drawing sheets and other documentation involving 10,000 pages. These drawings are stored by number and are actually completely inaccessible for the worker who usually asks the question of where and when a similar design has already been developed or in what construction a particular component has previously been used in combination with some other particular component. In certain design bureaus the working drawings are destroyed after a few months, since within the limits of this time period the so-called design memory operates, permitting retrieval of the required drawing among the other stored drawings. When drawings are accumulated over several years such searches are impossible. As for the museums, here thanks to the efforts of scholars, historians, archeologists, and regional scholars such a large number of specimens has been collected that the majority cannot be located in expositions and are stored so that they are not accessible for viewing. The question arises of how to create an information system which, for example, will enable each archeological find to become immediately accessible to all other archeologists, and so that the visitor to one museum would have access to the exhibits of all other museums [27].

The third form of documental information is economic (production, commercial, transport, public service, etc.). The rapid growth of science and engineering entails an expansion of the old industrial enter-

prises and the appearance of new industrial giants. In these enterprises use is made of hundreds of forms of machines, consisting of many thousands of varieties of standard components and parts, and hundreds of production items are produced. The planning section of such an enterprise must carry out periodic reviews of the production plans and operating programs, determine the actual volume of the product released for production, determine the required amount of raw material, calculate the proposed departmental loading, follow the progress of the technological process and the condition of the machinery, keep track of the operating documentation, and process statistically the data obtained. The activity of the supply and sales divisions must be correlated precisely with this operation, without which smooth operation of the enterprise is impossible.

The dense network of commercial communications, which long ago acquired an international nature, creates at the points of intersection – reshipment points (marine, river and airports, highway and railway junctions, warehouses and storage areas) – accumulations of the most varied cargoes which require accounting and distribution. The information operations in transportation include the recording of transit cargoes, integration of cargoes, and also the purely dispatcher functions – the compiling and monitoring of schedules with efficient utilization of time, transport routes and facilities. Under the conditions in large stores correctly presented information makes it possible to account for user demand, provide uninterrupted supply, and prevent the accumulation of slow-moving goods [2].

There is still another form, normally termed business documentation, which is closely allied with economic documentation, but has a broader nature. It includes primarily correspondence and various administrative and legal papers: orders, instructions, decisions, etc. The

old familiar expression "red tape" at first glance is directed against paper – the information carrier, just as we once heard the cry "Down with the machine!" Actually, at the moment we must talk about the destruction not of papers themselves but rather the old-fashioned methods of handling paperwork, we must place the operation of the office on the same level as that of the automated shop.

Organization of Mental Work

The problem of documental information is closely allied with the general problem of the organization of mental work. The fact that the rate of increase of the number of managing, office, and scientific personnel exceeds the rate of increase of the world's population [4] is indicative of the unsatisfactory organization of important spheres of human activity. In essence all the efforts of the recent centuries have been directed toward increasing the productivity of labor, while the work of the engineer or the scientist remains practically unchanged. We write, read, talk, and seek solutions all 7 hours of the working day, thinking about the decisions we have made in the most unlikely surroundings – at a meeting, in the dining room, or on the bus. Only a single process of mental labor has been mechanized to an adequate degree. This is the process of calculating, which is now accomplished with the aid of adding machines, desk-top electromechanical computers, and powerful electronic units which permit, for example, after 12 years (1946-1957) reduce the time for calculating 707 decimal digits of the number π from 25 minutes to 25 seconds [1]. The time has come for the emancipation of mental labor, reducing the proportion of the subsidiary processes which accompany the primary process – thinking.

At the Case Institute of Technology in the USA the time (in %) of the scientific worker is distributed as follows [28]:

scientific information (contact for information purposes)	33.4
business information	10.4
thinking (planning)	6.0
equipment setup	6.2
use of equipment (conduct of experiment)	23.4
data reduction	6.4
personnel and community affairs	9.8
various	4.4

The quantity 33.4% is, apparently, typical, and only chemists, as a rule, expend more time on information (in the strategic branches up to 94%) [4]. On the average, the chemical engineer reads during his life 0.5% of these published specialized literature, and half of this quantity relates to information of a general nature and has no connection with the work he is performing [29]. A colorful description, relating equally well to the Soviet specialist, is found in the article of the famous French documentalist G. Cordonnier [17]: "Many engineers working in large enterprises and in institutions have the habit of keeping on their table a pile of journals which they never have enough time to read, meanwhile these same journals are awaited anxiously by other readers who are less "high" and therefore are located further down the reading list. When this pile becomes excessively bulky, it is taken care of decisively, glancing hurriedly in one journal, leafing through another, and as a result reading nothing." This situation exists even in spite of the fact that far from all results of investigations and discoveries are written up in the form of articles, and technical reports, written in a hurry, leave much to be desired with regard to quality.

These deficiencies are reflected in the extremely low productivity

of mental labor which in the aircraft and radio industry of the USA is estimated at 10% [31]. Therefore there is mass duplication of scientific studies and engineering decisions, as a result of which 30-85% of the working time is irretrievably lost [34]. The Department of Defense of the USA alone expends 1-2 billion dollars annually as a result of duplication. The principal causes of duplication are [31]: 1) secrecy; 2) difficulties in disseminating and collecting information (imperfection of the system); 3) limitations in information exchange as a result of competition and interdepartmental barriers; 4) deficiencies in the organization of the enterprise operation (internal information); 5) improper understanding of the problem (terminclogical and other difficulties); 6) governmental limitations on the exchange of information between countries.

Documentalistics - a New Branch of Cybernetics

The word "documentation" has a three-fold meaning. We may understand by this the process of compiling documents, the set of documents on a given question, and a specific scientific discipline. The latter (its synonym is documentary science) is devoted to the official and technical questions of operation with documents and formally includes such practical aspects as scientific and engineering information, archive science, library science, patent affairs, museum management, office clerical work, etc.

The element of documentation is the document - fixed information which may be used for consultation, study, and proof. It may be shown that at the present time every fact is enclosed by clothing consisting of documents [32], whose forms become extremely varied - these are books, newspapers, journals, photographs and movie films, magnetic and phonograph recordings, catalogs, advertisements, reports, letters, card

catalogs, museum and exhibition displays. The modern conditions of operation with documents require approaches which are new in principle. Therefore the introduction of the ideas and methods of cybernetics into documentation practice has led to the creation of a new scientific discipline - documentalistics. Its basis is one of the divisions of cybernetics - the theory of information systems. This trend is being developed at the Center for Studies in Documentation and Communication (USA), the Documentation Center of the National Institute of Scientific Research (France), the International Documentation Center in Stockholm, the International Federation on Documentation (the former International Institute of Documentation), the Documentation Institute of GDR, the VINITI and the Scientific Council on Cybernetics of the Academy of Sciences USSR. The investigation materials are published in "American Documentation," "Nachrichten fuer Dokumentation," "Dokumentation," "IRE Transactions on Engineering Management," "Nauchno-tekhnicheskaya informatsiya," "Zavodskaya laboratoriya," and others. There is a gradual recognition of the fact that without cybernetics - the science of control and organization - the value of machines in the field of productivity of mental work cannot in itself be high [10].

A new practical profession has appeared - documentalist [14, 26, 33, 34], who in contrast with the librarian or archivist is first of all a specialist in that field of knowledge which the information service deals with, who actively utilizes such forms of documentation as bound clippings and photocopies, punched cards and microreproductions. The documentalist is faced with broad problems, which can be formulated in 10 basic points:

1) furnish the user all required information;

2) furnish information from adjacent fields which are of indirect interest;

3) provide timely information, i.e., not when the need for the information has already passed or has not yet arisen;

4) and in the shortest possible form;

5) provide additional information without any effort on the part of the user;

6) indicate the level of credibility;

7) know the source;

8) report possible delays in obtaining the information;

9) do not provide unnecessary information (combat redundancy);

10) aid the specialist in formulating his scientific and engineering reports by extensive use of the survey forms of documentation.

The reorganization of the information service in order to satisfy these conditions is impossible without special preparation of personnel and special assignments. During the years of the appearance and rapid development of documentalistics in the USA the average amount assigned for scientific information increased from 2% (1954) to 5% (1959) and, according to the experts, will increase again by a factor of 2.5 [35, 36]. The average number of specialists per information worker in 1956-1958 was 54 [37] and has a tendency to decrease. In the opinion of the Soviet researchers, the optimal level of expenditures on information search should amount to 1/5 of the expenditures on science [38].

Theory of Documental Information Systems

The information service in the human society is a complex of simple and complex information systems, each of which includes such stages of the information process as information collection, primary processing, transmission, secondary processing, storage, search and selection, correlation, duplication, and distribution. With respect to objective the systems are divided into reference, current accounting, diagnostic,

correlation, educational, game, etc., to each of which the fundamental cybernetic parameters are applied – operating speed, reliability, adaptivity, life, compactness, economy [3].

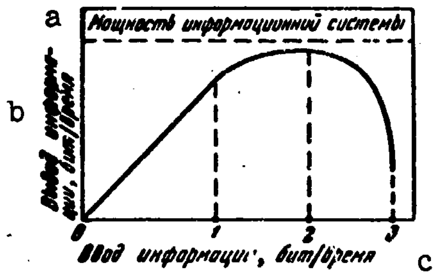


Fig. 3. Overload of information systems. 1) Region of normal operation; 2) overload region; 3) failure region. a) Capacity of information system; b) information output, bits/time; c) information input, bits/time.

The systems are intended for encompassing information flows and arrays which are characterized by the capacity and dynamics. Insufficient operating speed may lead to overloading and, as a result, to reduction of the operational liability leading to temporary and complete failure (Fig. 3). The other system parameters are related in the same fashion. For example, with limitation of the mechanical speeds we can increase the actual operating speed at the expense of compactness, etc.

The element of documental information is the document, which contains a definite number of information units (bits). The information object in the systems, other than documents, may be people, enterprises, materials, equipment, products, etc. Each of the objects is characterized by a definite set of information labels, among which we identify the so-called search labels. Taken together, they form the search image of the document, enabling us to find rapidly the required document in a large set of objects. The set of all search labels of a system is termed the search matrix, for which the number of cells is in a direct relationship with the number of objects. The term "density" denotes the number of cells per unit volume or area of the matrix. The

matrix operates with a binary code, i.e., each cell is assigned a definite label and relative to each object the cell responds to an interrogation by "yes" (1) or "no" (0). With this very simple arrangement of the system (direct coding), when all the labels are parallel and may be encountered in the objects in all possible combinations, the information capacity of the object matrix and of the system as a whole is the same. Mutually exclusive labels within the limits of a single aspect permit the use of combination coding, when the search label is assigned to a definite combination of cells; in this case the information capacity of the system increases sharply, while the capacity of the object matrix diminishes correspondingly (up to 1). In a demographic system (where the information object is the human being) an example of aspects with mutually exclusive labels are the surname, given name, patronymic, year of birth, nationality, residence. With combination coding use is made of different keys which facilitate the formulation and operation of the system.

Systems for current accounting are characterized by the fact that the number of information objects increases more rapidly than the number of labels. In the reference systems this increase is approximately the same, and a set of stereotype schemes is suggested for the search. The diagnostic systems (task: identify an unknown object with a set of known objects) are a variety of the reference systems, with the difference that hear the search strategy is essentially a sequential task. The correlation systems do not have the objective of searching out a required object, but permit performing a broad comparison of various labels in order to correlate them.

Sometimes the information objects and the labels exchange places. In this case we speak of the label search matrices and term such systems "inverted." An example are the compact card catalogs of archive

materials, where the card is entered not by document but by label and fixes the number of the corresponding documents.

From the point of view of mathematics any information system is a mathematical table with a certain number of inputs – from one (single-aspect systems) to several tens and hundreds. In the case of the single-aspect system we have an infinitely long tape which may be cut into individual "frames" (objects) and a conventional card catalog may be compiled. The "frames" are arranged in a strict order in the tape and in the card catalog. The only difference is that they cannot be removed from the tape and, consequently, it is difficult to supplement and renew them. With the card principle we can add to the system and select a search scheme such that we do not need to pass the entire array through the search device, but in this case the danger of disappearance of information increases markedly, so that much time and attention is required not only to find the card but also to put it in its place; in large arrays the incorrect return of a card is equivalent to its destruction. An example of a table with two inputs are the conventional (traditional) tables consisting of rows and columns, in which finding the required cells is accomplished by the coordinate method. The design of multi-input tables (multi-aspect systems) has some fundamental characteristic features and is based on the fact that the input function is performed by all the cells of the table (matrix). Thus, the card, or the tape "frame" becomes a point of a multi-dimensional information space, developed onto a plane [4], and becomes a part of the technically realized information system. At the present time there are punch, magnetic, and optical methods of realizing such systems.

The study of the comparative efficiency of the documental systems shows that the difference between the number of objects and the search labels must not exceed two orders. So far schemes have been developed

only for simple systems with a number of objects up to 10,000 and a number of labels up to 100 or a little more. The operation of a system with millions of objects requires the use of 10,000 labels, operation with which encounters difficulties in the classification aspect. The way out of this situation may be the compilation of the so-called descriptor glossaries and the development of local decimal classifications, LDK, which have quite flexible connection with one another within the framework of relatively large information services. The use of the universal classifications (of the UDK type) is increasingly considered inefficient [34] for the following basic reasons: such classifications are constructed using the differentiated principle and do not reflect the process of integration in science; the new and progressive branches of knowledge acquire a longer classification index than the older, traditional branches; the processes of developing classifications and their utilization are associated with major difficulties which make the information system unreliable and nonadaptive, providing a poor answer to today's problems.

So far we have discussed only the search labels whose objective is to provide for the retrieval of the required information. When the search is accomplished the system refers the user to a separately located array of documents arranged by number. This method of search is termed two-stage. Sometimes it is economically advantageous to introduce the uncoded information directly into the search system in order that the search operations deliver immediately the required documents or their copies. In this case the problems are solved not only speedily but also compactly, since there is no need to utilize extensive libraries and archives.

Mechanization and Automation of Information Systems

The tasks of integrated mechanization of information systems include operational search on the basis of the punch, magnetic, and optical forms of search matrices, the creation of magnetic matrices with high density of code storage (large electronic computer memory), the microreproduction of documents for compact storage, fast duplication and readout, provision for inexpensive machine translation of text from one language to another, the introduction of new methods for transmitting documental information, etc.

The punch forms of matrices have the lowest code storage density and are the cheapest. Here the basis is paper (cardboard) in the form of punched card or punched tape, which have their advantages and disadvantages, mentioned above, and which are quite widely used at the present time [40, 41]. However the punch card method is more highly developed and, apparently, is receiving general recognition for use in simple, nonautomated systems.

With respect to method of coding the perforated cards are divided into punched and notched. In the first case the perforation openings of the matrix are punched in the coding process; in the second case they are made ahead of time and are notched out during coding (the notches are completed to the edge or are connected with one another). With respect to location of the openings we differentiate cards with internal and external (edge) perforations.

The punched cards with internal perforations have been widely used as the so-called machine punch cards [2, 42-45] of standard format, 187.4 × 82.5 mm, whose matrix consists of 80 columns with 10 basic (0, 1, 2, 3, 4, 5, 6, 7, 8, 9) and 2 additional cells in each column, which constitutes in all 960 cells. In addition to these 80-column cards, use is made (considerably more rarely) of 90-column, 45-column, etc. The

perforators used to punch the required openings are designed similar to typewriters and operate with approximately the same speed, which as the result of a memory unit may be increased when duplicating perforations of the same type [46, 47]. A specialized reproduction device provides an output of 100-120 cards/min [49]. Recently portable (pocket) types of perforators operating on the principle of the automatic pencil [48] have come into widespread use. Sorting of the punched cards is accomplished at a rate of 25,000-60,000 cards/hr [46]. A special card selector processes simultaneously two arrays of cards at a rate of 250-650 cards/min [49]. In addition to these devices, the complex of the perforating machines include special perforators operating from pencil marks, verifiers, tabulators, decoders (60-65 cards/min), and card counters (100 cards/min) [47, 49].

The notched cards do not require complex mechanical preparation: the working tools are quite simple - scissors for cutting out openings and knitting needles for selecting the required cards [50]. The modest technical equipment, usually not exceeding in complexity a typewriter - shchiptsy-komposter, notching machines, electro-vibration selectors, etc. - provide some time saving in coding and multi-aspect searches, which require simultaneous use of several matrix cells. The advantages of the edge-punched cards are: simplicity and clarity of operation, the presence of a large field for uncoded information, the possibility of manual preparation. The advantages of the cards with internal perforations (they are also termed slotted cards - from the dominating form of cutout) are: somewhat greater capacity of the matrix, and the possibility of performing multi-aspect search in a single pass.

A special type are the so-called transparent, superposition punched cards, which usually operate using the inverse principle: the card is set up on the basis of a single label and each cell of the ma-

trix corresponds to a single object; superposing one card on another leads to alignment of the openings corresponding to objects with the given set of labels [51]. Recently the combined types of punched cards have become popular: machine cards used as transparent [52, 53], slotted cards used as transparent [54], transparent with edge perforation [55, 56].

The diagram of the relationship of the various types of punched card systems with respect to number of objects and labels (Fig. 4) shows that they compete very little with one another and, as a rule, mutually supplement one another. Cases are known when the rapid growth of the information array led to replacement of manual punched cards by machine cards. However, the Powers Regulator Co. in the USA considered it advisable to go the other way and converted a bibliographic reference system operating electronically with 80-column cards to cards with edge perforation in order to use more efficiently the uncoded (abstract) material; an important conclusion was drawn at the same time that it is not always nor in all cases wise to use digital computers and specialized servicing personnel [57]. The sorting speed for punched cards with edge perforation is usually 1/4 to 1/2 the sorting speed of the machine cards, and with arrays up to 100,000 units the manual sorting may be preferable [58]. Of considerable importance here is a factor such as the rate of usage (loading) of the system [59]. Stanford Research Institute in the USA carried out special experiments which showed that for a volume of the information flow of 40,000 documents per month the manual punched cards are preferable if the number of searches per month does not exceed 250; with 1000 searches per month the machines justify themselves if the volume of the information flow does not exceed 7-8000 (Fig. 5).

To date the optical forms of matrices do not have great practical

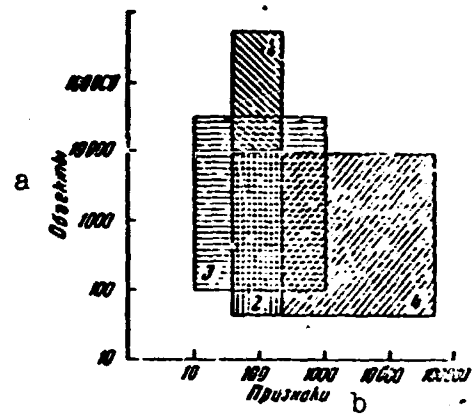


Fig. 4. Relationship of the various types of punched card systems with respect to number of objects and labels. 1) Machine punched cards; 2) transparent punched cards; 3) punched cards with edge and internal perforations; 4) transparent punched cards with edge perforations. a) Objects; b) labels.

importance. They are combinations of black (nontransparent) and white (transparent) points on a photographic surface occupying half or one-third the area of microcards. In the case of microfilms the matrix is photographed along the edges (in the manner of a sound track) or in the spaces between the frames of the microtext. The advantages of the optical method are: high density of code storage and long life of the coded information. The drawbacks are: relative expense and the technical difficulties associated with coding and recoding. It appears that these shortcomings may be avoided by the method of thermoplastic storage, which provides a density of 6.2 million cells/cm^2 [3].

An example of a fast but very expensive system is Minicard, where a 16 × 32 mm microcard contains the image of 12 pages of text and matrices with 2730 cells. Documentation consisting of 11 million pages is stored on 900,000 cards. The system operating speed is 1800 cards/min.

Microfilm devices with revolving endless tape (of the type of Rapid Access Look-Up, Poisk, etc.) have limited application because of the inadequate number of information objects and the inadequate volume of the matrix, and also because of the difficulties associated with addi-

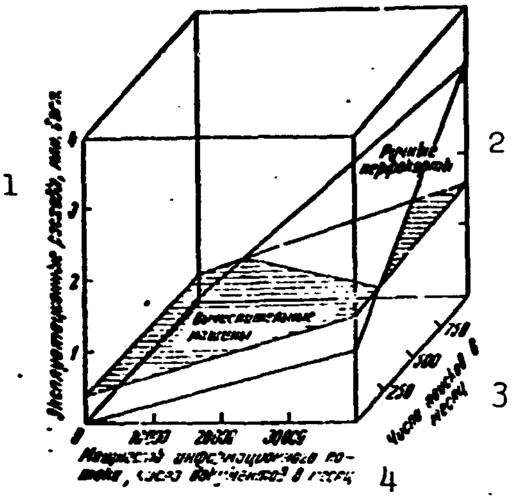


Fig. 5. Economics of machine and manual (punch cards with edge perforation) information systems. 1) Operational costs, millions of dollars; 2) manual punch cards; 3) number of searches per month; 4) volume of information flow, number of documents per month.

tion to and renewal of the documental information. As a rule they are used in commerce and reproduce commercial reference materials, providing the capability of finding the required page in 1.5 seconds.

The magnetic method has begun to play a major role in the electronic information-logic machines and has far from exhausted its capabilities, particularly with relation to magnetic cards which contain, like the optical cards, a microimage of the pages of the documents and a magnetic matrix (along the card edges or in a block at one side — in the first case the microimage is read in transmitted light, in the second case it is read in reflected light). The advantages of this method are: comparatively low cost, very high recording density, and the possibility of systematic rewriting of the code. As a result of the primary shortcoming — unreliability of the storage — the electronic information-logic systems using the magnetic form of memory are forced to store the primary information material on machine cards which systematically copy the object matrices onto the magnetic base.

As an example of a system using the magnetic method, we mention

Magnavue (the firm Magnavox). One-third the area of the 25.4 × 76.2 mm card is occupied by a microphotocopy of the text, the remaining part of the card is occupied by a matrix of capacity 3000 cells. The minimal service life of the card is 200,000 scannings. The information input rate into the computer is 90,000 symbols/sec [3].

The struggle for high speed and large memory of the information systems is not limited to the methods just mentioned. In the USA in the near future information stores of capacity 10-100 billion cells will be built, and the cost per cell will be reduced to 5 cents using magnetic drums and discs [3]. Realistic conditions are being created for storing with the aid of electrical impulses on a thin metallic film coating a glass plate of dimensions 13 × 18 cm the equivalent of the information stored by the human memory [50]. We predict a great future for the crystal optics forms of memory, where the information cells are the colored defects of the crystal lattice.

Along with these directions, problems are being resolved of reducing (reduction of size without alteration of form) uncoded information. Today's microphotoreproduction techniques permit obtaining efficiently microcopies of documents reduced by a factor of 10-20, permits reading them with the aid of desk-top and pocket readers, and permits making any number of copies of normal dimensions [60-63]. The cost of the microcards is from 5 to 50% of the cost of the books which they reproduce [14]. The microcards carry a magnetic or optical form of search matrices [3] or are combined with punched cards (aperture and klyassernyy cards) [2, 34, 64-68], which are, therefore, an element of relatively optimal (cybernetic) information systems. The Verac system with the aid of 140-fold reduction permits reproducing on 25.4 × 25.4 cm film 10,000 pages of text; the information stored on 200 such films is searched in 2 seconds and then is transmitted over television, coded on

paper in normal scale or on photofilm with a standard reduction [3]. The Walnut system (IBM firm) in 2 passes performs a reduction to 1/1000 of the natural size. In each of 100 modules there is stored 200 plastic containers with 50 microcards. The search is performed using punched card, punched tape, or magnetic methods [3]. In principle it is possible (but not yet technically accomplished) to reproduce on a sphere the size of a pin head the text which is equivalent in volume to 24 volumes of the Encyclopedia Britannica. The required information will be found by the coordinate method and read with the aid of an electron microscope.

Thus, all the books in the world may be concentrated in the volume of a pocket notebook. This tendency to bring the information near the user is of great importance in selecting a technical policy in the field of documental information.

Organization of the Information Services

The ideas of documentalistics find their reflection (elementary for the moment) in the practice of the information services. Library [69-71] and archive [72-73], including drawings [3], materials, current newspaper and journal information [74-75] are being translated onto microfilm. The world's first scientific journal on microfilm, bypassing the conventional method of printing, has begun publication and its primary advantages are: low cost, quick publication, retention of articles of large volume, improved quality of illustrations, convenience of storage [76]. A basically new form of publication is the so-called electronic book – "Computexts" – a magnetic tape suitable for input into a high-speed digital computer. The first such issue is entitled "Electronic Index of Chemical Patents" (150,000 descriptions covering 17 years, volume 0.028 m^3) [77]. The American Society of Chemical En-

gineers has made it mandatory to provide the articles published in its organs with descriptor abstracts which are suitable for entry into any information system, and has issued in two volumes a specialized descriptor dictionary (the source) on chemistry and chemical engineering (50,000 terms) [78]. An example of efficient organization of library servicing is the storing of documents on aperture punched cards and their transmission over specialized television channels within a radius of several kilometers. The client contacts the book depository by telephone and at an agreed time switches on the television, equipped with an automatic focusing system and adaptors for turning the pages, changing the image area and the magnification ratio [79]. The highly automated electronic systems are still of an experimental nature. They include the information service of the Central Information Agency of the USA, which operates on the Walnut system and within 10 seconds finds any document from 100 million stored documents [80]. The electronic machine Sinoptikon with a printer providing 600 lines/min compiles a bibliography on a given subject and writes out citations to questions — "family," "man," "god," "punishment," and others from the books of 74 authors and the Bible [81].

The experience accumulated makes it possible to draw certain preliminary conclusions. First, the selection of limited-mechanization or highly mechanized (automated) systems is determined primarily by two factors: the intensity of system usage and the nature of the information requirement. In purely reference systems, which produce finished results, the level of mechanization is dictated only by the intensity of utilization: an expensive system which services a client in one second will be preferable to a cheap system which does the same in one hour, if we must satisfy hundreds of requests per day. In bibliographic and other such systems which provide data in semi-finished form for

further individual processing, limited mechanization is required (at each writing table) or a combined system which includes limited mechanization in the last stage.

As an example we cite the optimal organization of a system for medical diagnostics and correlation of medical (clinical) experience. Each practicing doctor maintains his own personal manual punch-card catalog of disease histories. The medical information center collects duplicates of these cards, reproduces them and sends them on a subscription basis to all interested individuals and organizations, who enter them into their own narrowly specialized information systems. In turn the machine duplicates of the same cards are input to a single reference automated system whose primary purpose is to establish precedents and select data from them, the development of optimal methods of diagnostics, prophylactics, and treatment of diseases.

For the first time VINITI in 1963 began to publish the abstract journal "Metallurgiya" on manual punched cards, which permits a sharp increase of the number of individual subscribers to the abstract volumes, increases the profitability and efficiency of these volumes. Metallurgiya appears in 23 individual card catalogs, coded, in addition to the library-bibliographic labels (author, publication type, branch, country, card file number, and year), by 100 basic (constructed using the LDK) and 40 additional subject labels (headings), 3000 headings in all. The specialized information centers, obtaining information in the form of such punched cards, recode the information (without altering the form of the stencil) in accordance with its local decimal classification (10-100 card files with 100 or 140 headings in each), circulate the information according to the requests of their users (laboratories, factories and individual specialists), and send the cards to locations where they are recoded a third time and introduced into the specialized

local information systems. One of the ten card files on general questions of a given field is used directly in the center itself for the local 3rd stage system. The reference service of this center involves the coordination of the operations of all the highly specialized systems and getting them to form regular reference functions on request through the center.

This three-stage information system [82] facilitates the work of the information services by separating the functions of information collection and processing, replaces the tedious universal classifications by adaptive local schemes which are consonant with the structure of today's economy, permits better utilization of the information materials.

The issuance of secondary information on punched cards may be considered as the operation of merging two information processes: abstracting and searching. The next step is the merging of the primary and secondary information so that the issuance of each publication would simultaneously input it into the existing search systems [34]. This task may be performed by punched cards or other new forms of cybernetic technology.

REFERENCES

1. A. Ducrocq. L'ere du calculateur electronique et du traitement de l'information.[The Age of the Electronic Computer and Data Processing], Procedes electron, 1961, No. 9, 6.
2. Perforirovannyye karty i ikh primeneniye v nauke i tekhnike [Punch Cards and Their Application in Science and Engineering], Mashgiz, 1963.
3. Bushor. Khaneniye i otyskaniye informatsii. [Information Storage and Retrieval], Elektronika, 1962, No. 26, 15.

4. G.E. leduts, V.V., Nalimov, N.I. Styazhkin. Nauchneya i tekhnicheskaya informatsiya kak odna iz zadach kibernetiki [Scientific and Engineering Information as a Problem of Cybernetics], UFN, 1959, 60, No. 1, 13.

5. W.H. Huff. Indexing, Abstracting, and Translation Services. Libr. trends, 1962, 10, No. 3, 427.

6. Scientific data "rotting in storage." – Aerospace Manag., 1961, 4, No. 10, 74.

7. M. Kherner, S. Kherner, Otchety o nauchnykh issledovaniyakh, provodnmykh pravitel'stvennymi uchrezhdeniyami Soyedinennykh Shtatov Ameriki. Sovremennoye sostoyaniye voprosa. ["Reports on Scientific Studies Performed by Governmental Activities of the USA. Present State of the Question."] Byull. YuNESKO dlya bibl., 1959, 13, 8-9,203.

8. A.I. Mikhaylov. Tekhnicheskiy progress i zadachi nauchnoy informatsin ["Technical Progress and the Problems of Scientific Information"]. Vestn. AN SSSR, 1962, No. 2, 29.

9. Rost vypuska khimicheskoy literatury ["Growth of Published Chemical Literature"]. Inf. byull. zapubezhn. khim. prom. 1962, No. 16 (119), 63.

10. D.A. Brown. Agricultural Periodicals. Libr. trends, 1962, 10, No. 3, 405.

11. N.N. Kukharkov. Gosudarstvennaya registratsiya proizvedeniy pechati. [State Registration of Press Output], Collection: "Voprosy organizatsin i metodiki nauchno-tekhnicheskoy informatsin i propagandy" [Questions of Organization and Techniques of Scientific-Engineering Information and Dissemination]. Mowcow, 1960, 60.

12. Koblans, G., Novyye metody i tekhnika soobshcheniya sved-

niy [New Methods and Techniques for Communicating Information], YuNESKO dlya bibl., 1957, 11, No. 7, 154.

13. Brown, L., Future of Scientific Paper. J. Docum., 1962, 18, N 1, 1.

14. Farber, E.I., General Periodicals. Libr. trends, 10, No. 3, 310-320 (1962).

15. Welch, H.M., Economics of Periodical Publishing. Libr. trends, 1962, 10, No. 3. 293-301.

16. Fozzy, P., Research coordination. Bull. Atomic scientists, 1962, 17, No. 3, 29.

17. Kordon'ye, Zh. , Primeneniye mikrokopirovaniya v praktike tsentrov dokumentatsii [Use of Microcopying in the Practice of Documentation Centers]. Byull. YuNESKO dlya bibl., 1958, 12, No. 11-12, 301.

18. Kofnovec, L., Veralten wissenschaftlich technische Information [Obsolescence of Scientific and Technical Information]. Dokumentation, No. 2, 42.

19. Feinstein, P.S., Translation Activities of the Foreign Science Information Program. Spec. libr., 1962, 53, N 1, 26.

20. Mikhaylov, A.I.. Mezhdunarodnaya konferentsiya po nauchnoy informatsii [International Conference on Scientific Information]. Vestn. AN SSSR, 1959, No. 6, 103.

21. Price, J. De S., Science Babylon. Yale Univ. Press, New Haven, 1961.

22. Information Methods of Research Workers in the Social Sciences. Ed. Harvey J.M. London, 1961.

23. Masanov, Yu.I. Za dal'neysheye uluchsheniye raboty knizhnykh palat [A Call for Improved Operation of Publishing Houses], Sov. bibliografiya, 1957, No. 45, 3.

24. Nuzhdy spetsial'nykh bibliotek [The Needs of Specialized Libraries], Byull. YuNESKO dlya bibl., 1958, 12, No. 11-12, 282.

25. Oznakomleniye amerikanskikh khimikov s nauchno-tekhnicheskoy literaturoy [Acquainting American Chemists with the Scientific and Engineering Literature], Inf. byull. zarubezhi. khim. prom., 1962, No. 16 (119) 63.

26. Vorob'yev, G.G., Zadachi kibernetiki v oblasti arkhivnoy dokumentatsii [Problems of Cybernetics in the Field of Archive Documentation]. Vopr. arkhivoved, 1963, 8, No. 4, 94.

27. Vorob'yev, G.G., Nekotoryye voprosy primeneniya kibernetiki v arkheologii [Some Questions of the Application of Cybernetics in Archaeology], collection: "Metody yestestvennykh i tekhnicheskikh nauk v arkheologii" [Methods of the Natural and Technical Sciences in Archaeology], Izd-vo AN SSSR, 1963, 60.

28. Halbert, M.H., Ackoff, R.L., An Operations Research Study of the Dissemination of Scientific Information. Intern. conf. scientif. inform., Washington, 1959, 88.

29. Case Institute Study of Scientific Communication Among Chemists. Science inform. news, 1959, 1, N 1, 8.

30. Voigt, M.J., Researcher and His Sources of Scientific Information. Libri (Denmark), 1959, 9, N 3, 177.

31. Hirsch, I., et al., Relation of Utilization the Shortage of Scientists. IRE Trans. Engng Manag., 1958, EM-5, No. 3, 73.

32. Briet, S., Qu'est-ce que la documentation? [What is Documentation?] Paris, 1951.

33. Rupprecht, H., Dokumentalist und Informationsingenieur – neue Berife. Ein Beitrag zur Ausbildung von Dokumentations – und Informationskräften in der DDR [Documentalist and Information Engineer – New Job Titles. A Contribution to the Training of Documentation and Information Workers in the GDR]. Technik, 1961, 16, N 8, 589.

34. Nalimov, V.V., Vorob'yov, G.G., Kibernetika v koordinatsii nauchno-issledovatel'skikh rabot [Cybernetics in the Coordination of Scientific Research Operations], Zavodskaya laboratoriya, 1963, No. 3, 259.

35. Kent, A., Perry, J., Centralized Information Services (opportunities and Problems), N.Y., London, 1958.

36. Spitzer, E.F., McKenna, F.E., in book. "Information and Communication Practice in INdustry" N.Y., 1958.

37. Hanson, C.W., Survey of Information (Library Units in Industrial and Commercial Organizations). ASLIB processing, 1960, 12, No. 11, 39.

38. Arutyunov. Vazhnyy etap v razvitii nauchno-tekhnicheskoy informatsii [An Important Step in the Development of Scientific and Engineering Information], Ekonom, gazeta, 30 June 1962.

39. Knox, W.T., Informationsarbeit für die Forschung [Information Work for Research], Nachrichten für Dokumentation, 1962, 3, No. 1, 1.

40. Schroder, D.L., Are Punched Cards Becoming Obsolescent? J. Mach. Account, 1961, 12, No. 10, 13.

41. Dreyfack, R., Paper Tape? Punched Card? – Administr. Mag., 1961, 22, No. 11, 22.

42. Erste europäische Lochkartenmaschine. [The First European

Punchcard Machine], Bürotechn. Automat., 1961, 2, N 8, 230.

43. Gall, M.W., Informationsverarbeitung. [Information Processing] Elektro-Jahr 1962, Wurzburg, 1962.

44. Both, A.D., Machines that Think. J. Roy. Soc. Atrs. 1962, 110, N 5069, 345.

45. Hadlich, H., Anwendung von Maschinenlochkarten bei der Dokumentationsarbeit. [Use of Machine Punchcards in Documentation Work], Forsch. Lehre-Prax., 1962, No. 4, 114.

46. Mater, E., Über die Verwendung von Maschinenlochkarten für wissenschaftliche Aufgaben [Use of Machine Punchcards for Scientific Purposes]. Forsch. Lehre-Prax., 1962, N 4, 80.

47 Mater, E., Funktionsweise von Lochkartenmaschinen [Operating Principle of Punchcard Machines]. Forsch. Lehre-Prax., 1962, N 4, 130.

48. Biringer, R.T., Port-a-punching for Data Processing. Data Process, 1961, 3, No. 9, 22.

49. Punched Card Systems. Survey. Part VIII. Automat. Data Process, 1961, 3, N 9, 33.

50. Vorob'yev, G.G., Tekhnika sostavleniya i rabota s kartotekami iz perfokart ruchnoy sortirovki [Techniques of Compiling and Working with Manual-sort Punch Card Files], Nauchno-tekhnicheskaya informatsiya, 1962, No. 10, 23.

51. Beauclair, W. de, Zur Dokumentation mit Sichtlochkarten [Superposition Punched-Card Documentation], Automatik, 1962, 7, N 9, 360.

52. Vleduts, G.E., Ispol'zovaniye dlya informatsionnogo poiska standartnykh perfokart v kachestve superpozitsionnykh [Use of Standard Punch Cards as Superposition Cards for Information Retrieval], Nauchno-tekhnicheskaya informatsiya,

1962, No. 5, 25.

53. Otbor periodicheskikh izdaniy pri pomoshchi vzaimonalozheniya perforirovannykh kartochek [Selection of Periodical Publications with the Aid of Superpositioning of Punched Cards], Byull. YuNESKO dlya bibl., 1958, 12, No. 8-9, 252.

54. Dreyer, K.A., Eine neues Variante des Sichtlochverfahrens [A New Variable in the Superposition Process], Fermelde-Prax., 1962, 39, No. 21, 909.

55. Richter, W., Worseck, M., Literaturdokumentation mit Hilfe der Kombination von Kerb und Sichtlochkarten [Literature Documentation with Combined Notch-and Punchcards]. Dokumentation, 1962, 9, N 2, 53.

56. Th. H. Zink. Novel cataloging system. Appl. Spectroscopy, 1961, 16, N 1, 22.

57. Axelrod, L.R., An Information Retrieval System for a Small Research Department. IRE Trans. Engng Manag., 1962, 9, N 2, 92.

58. Hirayama, K., Time required, cost and personnel for documentation. In collection:"Tenth Pacific Science Congress, Honolulu, Aug. 21st-Sept. 6th 1961." Progr. and abstr. of paper for the section of scient. inform.

59. Kolder, R., Rabota Mezhdunarodnoy konferentsii po vychislitel'noy tekhnike dlya obrabotki informatsii [Proceedings of International Conference on Computer Technology for Information Processing]. Byull. YuNESKO dlya bibl., 1959, 13, No. 10, 246.

60. Saringulyan, M.Kh. Mikroprintirovaniye [Microprinting], Sov. bibliografiya, 1957, No. 45, 114.

61. Kombinirovannyy apparat dlya mikrofil'mirovaniya i chteniya

mikrofil'mov [Combined Equipment for Microfilming and Reading Microfilms], Byull. YuNESKO dlya bibl., 1958, 12, No. 10, 258.

62. Nedorogoy apparat dlya chteniya mikrofil'mov [Inexpensive Equipment for Reading Microfilms]Byull. YuNESKO dlya bibl., 1959, 13, No. 4, 105.

63. Portativnyy apparat dlya chteniya mikrofil'mov [Portable Equipment for Reading Microfilms], Byull. YuNESKO dlya bibl., 1961, 15, No. 3, 187.

64. Luther, F., Language of Liliput. A Thesaurus from Users of Microfilm. Libr. J., 1961, 86, No. 5, 929, No. 13, 2425, No. 17, 3238.

65. Isacson, S., Industry faces the information challenge. Design News, 1962, 17, 58.

66. Filmsort aperture card. Microfilm products. Minnesota mining and manufacturing Co. Business automat., 1962, 7, No. 6, 19.

67. Lipkin, S.A., Vorob'yev, G.G., Fotograficheskiye metody nakopleniya laboratornoy informatsii [Photographic Methods of Storing Laboratory Information]. Zavodskaya laboratoriya, 1962, No. 10, 1221.

68. Lipkin, S.A., Vorob'yev, G.G., Primeneniye aperturnykh perfokart dlya informatsionnogo poiska. 1. Izgotovleniye i chteniye aperturnykh perfokart [Use of Aperture Punch Cards for Information Retrieval. 1. Preparation and Reading of Aperture Punch Cards], Nauchno-tekhnicheskaya informatsiya, 1963, No. 3, 40.

69. Svyshe 5000 arabskikh rukopisey na mikrofil'makh [More than 5000 Manuscripts on Microfilms], Byull. YuNESKO dlya bibl.

1958, 12, No. 8-9, 243.

70. Atomnaya informatsiya na mikrokartochkakh [Atomic Information on Microcards], Byull. YuNESKO dlya bibl., 1959, 13, No. 10, 263.

71. Mineaturfilm von Bibliotheksbeständen [Microfilming Library Catalogues], Mikrofilm, 1962, 12, No. 1, 6.

72. Reprodutsirovaniye dokumentov. Peredvizhnaya sluzhba YuNESKO dlya mikrofil'mirovaniya v Paname [Document Reproduction. Mobile UNESCO Service for Microfilming in Panama], Byull. YuNESKO dlya bibl., 1958, 12, No. 10, 256.

73. Angliyskiye sudebnyye otchety na mikrokartakh [English Legal Reports on Microcards], Byull. YuNESKO dlya bibl., 1959, 13, No. 11-12, 306.

74. Gazety v vide mikrofil'm [Newspapers in Microfilm Form], Byull. YuNESKO dlya bibl., 1958, 12, No. 10, 258.

75. Mikrofil'my gazet i zhurnalov [Microfilms of Newspapers and Magazines], Byull. YuNESKO dlya bibl., 1959, 13, No. 11-12, 306.

76. Pervyy nauchnyy zhurnal, izdavayemyy na mikroplenke [The First Scientific Journal Published on Microfilm], Byull. YuNESKO dlya bibl., 1959, 13, No. 5-6, 147.

77. Doc Inc's computext. J. Chem. docum., 1962, 2, N 3, 190.

78. Holm. B.E., Information retrieval – a solution – Chem engng. progr., 1961, 57, N 6, 73.

79. Microfilm televiser. Perspective, 1961, 3, N 4, 343.

80. CIA's film memory to store 990,000 pages. Control engng., 1961, 8, N 7, 23.

81. La biblioteka elektronika [The Electronic Library]. Schede perfor. e calcobo elettron., 1962, 8, No. 43, 29.

82. Vorob'yev, G.G., Perspektivy organizatsii sluzhby nauchno-tekhnicheskoy informatsii s primeneniyem perfokart [Prospects for Organizing Scientific and Engineering Information with the Use of Punch Cards], Collection: "Vsesoyuzi. nauchno-tekhn. soveshch. po primen. kart s krayev. perforatsiey. Tezisy dokladov." [All-Union Scientific and Engineering Meeting on Use of Edge-Punch Cards], Tallin, TsBTI, SNKh ESSR, 1964.

Manu script Page No.	[Transliterated Symbols]
1, 7	ВИНИТИ = VINITI = vsesoyuznyy institut nauchnoy i tekhnicheskoy informatsii = All-Union Institute of Scientific and Engineering Information
28	ЭЦВМ = ETsVM = Elektronnaya tsifrovaya vychislitel'naya mashina = electronic digital computer
21	ЛДК = LDK = lokal'naya desyatichnaya klassifikatsiya = = local decimal classification
21	УДК = UDK = universal'naya desyatichnaya klassifikatsiya = = universal decimal classification

MICROPHOTOREPRODUCTION METHODS OF REDUCING DOCUMENTAL INFORMATION

S.A. Lipkin

Microphotocopying — the production of reduced photographic copies of documents — enables the solution of several fundamental problems concerning the mechanization of information work. Microphotocopying is used as a basis for providing compact storage, creating conditions suitable for mechanization and automation of information search, and solving the question of efficient copying and duplication of documents. All this permits a new approach to the problems of publishing and library functions, creating more flexible and efficient systems for bringing information to the reader.

At the present time in the USSR there are in serial production several models of equipment and specialized photographic films for microphotocopying, which provide the possibility of the practical introduction of this process in many fields of work with documents: scientific and engineering information, design drawing documentation, archive management, etc.

However, in the interest of progress of this very promising field of technology a considerable expansion of the output of equipment and photographic materials is necessary.

Modern microreproduction technology is characterized by a tendency toward obtaining as great a reduction of the document size as possible, large-scale use of microphotocopying, the construction of systems for mechanizing and automating information search based on microphotocopies

of documents, the design of facilities for efficient copying and duplication of documents.

I. REDUCTION OF DOCUMENT DIMENSIONS (INFORMATION CONTRACTION)

The size, volume, and correspondingly the weight of documents are reduced during microphotocopying approximately inversely proportional to the square of the reduction factor of their linear dimensions. If a document of dimension 20 × 30 cm is reduced by microphotocopying only by a factor of 10 (which is elementary with the modern technical capabilities), then its dimensions on the microphotocopy become 2 × 3 cm, i.e., the area and approximately the volume are reduced by a factor of 100.

Let us denote: R is the reduction factor of the linear dimensions of the documents; V_m is the volume of the microphotocopies; V_d is the volume of the documents; k is a coefficient of proportionality which takes into account the relationship of the thicknesses of the document paper and the photographic film, the nature of the folding (presence of air spaces), etc. Then the equation which expresses the relation between the volume of the documents and the microphotocopies takes the form:

$$V_{\text{м}} = k \frac{V_{\text{д}}}{R^2} .$$

The same relation will hold between the weight of the documents and their photocopies, the area of the documents S_d and of the microphotocopies S_m, i.e.:

$$S_{\text{м}} = \frac{S_{\text{д}}}{R^2} ,$$

as in this case $k = 1$.

The effect which may be achieved by reducing the dimensions of documents by microphotocopying is obvious: quite realistic technologi-

cal capabilities are created for relieving the document concentration centers of the paper storage facilities without reducing at the same time the completeness of the information which these centers contain.

It should be noted that in microphotocopying not only is an absolutely exact content of the documents retained, but all their graphical peculiarities, which in certain cases is very important.

At the present time we use in practice a reduction of up to 20 times in microphotocopying. There are capabilities (photographic materials and equipment have been fabricated) for reductions of up to 40 times. In several systems for automating information search which have been developed in the USA use is made of photocopying with a reduction of the original documents by 60, 100, and even 200 times [1]. With a reduction of 200 times the microimage of the page will have dimensions of approximately 1 × 1.5 mm, and 2625 microimages fit on a microcard of format 75 × 125 mm (dimension of a standard bibliographic card).

By using high density of document spacing, we can very efficiently solve the problem of rapid search without resorting to complex devices. If a large group of documents, which may be selected on a common theme, are arranged on a single microcard, then the search problem reduces to finding the coordinates of the site of the required document on the microcard.

The problem of achieving high reduction factors is associated with the development of photographic materials and optics with high resolving powers. Between the resolving power N of the photographic system "lens-photographic material," used for microphotocopying, the reduction ratio R, and the minimal line width on the document d there exist the following relation:

$$\frac{d}{R} > \frac{1}{2N}.$$

i.e., the line width on the microphotocopy d/R must be equal to or greater than the minimal width of the line which the photographic system is capable of transmitting. The width of the lines of the various typographic fonts usually varies in the limits from 0.1 to 0.25 mm.

The selection of the optimal resolving power of the photographic systems "lens-photographic material" is illustrated by the following example: for a minimal width of a line on the document equal to 0.15 mm, when preparing microphotocopies with a 20-fold reduction it is necessary that the resolving power (without account for unavoidable losses) be greater than 70 lines/mm (from the preceding relation

$$N = \frac{R}{2d} = \frac{20}{2 \cdot 0,15} \approx 70).$$

We have been able, using the new Mikrat-positive photographic film with resolving power 470 lines/mm, produced by the Kazan branch of NIKFI, using Soviet-made optics, to obtain quite clear images of typographic text (line width 0.15 mm) with a reduction greater than 60-fold.

II. THE TECHNOLOGY OF LARGE-SCALE MICROPHOTOCOPYING

Current microreproduction technology is based on highly productive equipment permitting the transfer of large arrays of documents to microphotocopies in a short time. The microphotocopying of documents is usually performed in two stages: first microphotocopy negatives are prepared, from which positive copies are then printed. The negatives serve for the printing of the required number of positives, which are used as the working material for retrieval and reading of the documents. If necessary the exposed photographic film may be processed using the so-called reverse process, permitting the production of a positive image without subsequent printing on positive photographic film.

The reverse process of handling the photographic film is considerably more complex than the common negative-positive process, and the principal difficulty here is the reproduction of the microphotocopies, which, as a rule, is required.

With the availability of high-quality microphotocopies the storage of the original documents is unnecessary, since the complete and exact content of the documents always may be reproduced from the microphotocopies.

1. <u>Photographic Materials for Microphotocopying</u>

As the photographic materials for microphotocopying we use the specialized so-called Mikrat photographic films produced by the Soviet movie film industry, primarily at the V.V. Kuybyshev Chemical Plant in Kazan.

All types of Mikrat films have a noncombustible (triacetate) base. Prior to 1962 a single type of negative Mikrat photographic film Mikrat-200 was produced having a resolving power of 200 lines/mm. Serial production of the Mikrat-300 film began in 1962. At the end of 1963 experimental lots of two new Mikrat photographic films – Mikrat-300B and Mikrat-positive became available.

In contrast with Mikrat-300, the Mikrat-300B has increased light sensitivity and a dark antihalo layer. This photographic film is intended for micro filming equipment for dynamic pictures.

The Mikrat-positive photographic film is used for contact printing of positive microphotocopies. It is also produced with a special anticurl layer, providing flatness of the photographic film, which is extremely important in the preparation of microcards (format pieces of photographic film with microphotocopies of the documents arranged on

them).

The following are the characteristics of certain types of Mikrat photographic films.

Characteristics of Mikrat Films

Тип фотопленки 1	Изготовитель 2	Дата выпуска 3	Номер эмульсии 4	Характеристика фотопленки 5: Светочувствительность, ед. ГОСТ 6	Контрастность 7	Разрешающая способность, лин/мм 8	Максимальная плотность 9	Толщина эмульсионного слоя, мк 10
Микрат-200 11	12 Фабрика № 5	13 Июнь 1959 г.	34	4,5	3,3	196	3	—
Микрат-300 11	14 Химзавод им. Куйбышева, г. Казань	15 Июль 1963 г.	15016	2,2	5,3	317	3	—
11 Микрат-300В	16 То же	17 Декабрь 1963 г.	15866	8,5	4,2	317	3	6,5
18 Микрат-позитив	»	17 Декабрь 1963 г.	15888	0,045	3	470	3	4,6

1) Type of photographic film; 2) producer; 3) production date; 4) emulsion number; 5) characteristic of photographic film; 6) light sensitivity, units per GOST; 7) contrast; 8) resolving power, lines/mm; 9) maximal density; 10) emulsion layer thickness, microns; 11) Mikrat; 12) plant; 13) June; 14) Kuybyshev Chemical Plant in Kazan; 15) July; 16) same; 17) December; 18) Mikrat-positive.

The Soviet Mikrat photographic films are equivalent in quality to the best foreign photographic films for the same purpose.

Diazo film. In contrast with the Mikrat photographic films described above which contain in the emulsion silver halide which is sensitive to light, the light sensitivity of the diazo films is the result of complex organic substances – diazo compounds, which decompose under the action of light [2, 3]. The diazo films have very low light sensitivity in the ultraviolet portion of the spectrum, therefore, they may be processed in a light room. The diazo films have high resolving power, exceeding the resolving power of the best Mikrat photographic films and have high contrast.

Wet processes, associated with subsequent long-duration drying,

are not required for developing the diazo films. Their cost is significantly lower than the cost of the silver halide films. These characteristics make the diazo films very promising for preparing positive microphotocopies. This method may be used for efficient duplication of documents directly at the storage location.

Kalvar photographic film. The light sensitivity of the Kalvar photographic film is also the result of diazo compound content [4]. The nitrogen atoms which are released during their decomposition under the action of light remain within the layer; after a slight short-term heating the layer softens and the nitrogen bubbles depart, leaving a trace (cavity). The light refraction coefficients are different on the segments of the layer where the cavities have been formed and where they have not. Depending on the nature of the illumination a negative or positive image is obtained. The resolving power of the Kalvar photographic film is 1000 lines/mm.

Photochromatic materials. The photochromatic materials. The photochromatic materials permit obtaining microphotocopies with very large reduction ratios. In comparison with the silver halides, the photochromatic materials have several important advantages: there is no graininess at all, a very high reduction ratio is provided, very high resolving power, the capability of transferring from small to very high densities, the image becomes visible immediately after exposure and does not require development, the resulting image is "reversible" - it may be easily removed and again impressed on the same segment of the photographic layer.

The production of the image on the photochromatic materials is based on the property of these materials to alter their color when illuminated by light of the corresponding portion of the spectrum; if then the resulting image is again illuminated by light of a different

portion of the spectrum or the light sensitive layer is heated slightly the image disappears. The reversibility of the process, the removal and repeated exposures on the same segment of the layer, is based on this property. Along with these valuable qualities the photochromatic materials have an essential shortcoming — poor retention of the image: under room temperature conditions the image is retained only a few hours, and at low temperature a few months or even a year.

2. Equipment and Technology for Mass Production of Roll Microphotocopies

The technological process of the production of microphotocopies is made up of the operations of photographing the documents, development, inspection and mounting of the negatives, printing and development, inspection and mounting of the positives.

Photographing of the documents is performed on specialized microfilming equipment. There are two types of such equipment: static filming, when the filming is accomplished with a stationary original (document) and stationary photomaterial, and dynamic (slit) filming, accomplished with continuous and synchronous motion of the original and the photographic material.

Figure 1 shows a general view of the UDM-2 static type microfilming apparatus produced in series by the Odessa Plant of KINAP. The documents in the form of individual sheets or bound in a book are placed under the glass on the illuminated screen 1, the plate 2 presses the surface of the pages against the glass. The light from the lamp 3 is reflected from the surface of the page being photographed, passes through the lens of the camera 4, and strikes the light-sensitive layer of the photographic film.

The UDM-2 unit uses roll perforated or unperforated photographic

film of either 16 or 35 mm width. The camera magazine is designed to hold up to 30 m of photographic film. The reduction ratio during microreproduction may be varied smoothly from 5 to 20. Focusing during change of the reduction ratio is accomplished automatically. The maximum format of the document being photographed is 407 × 576 mm. The setup is adapted for semiautomatic operation (the operator simply turns the pages); the productivity depends on the nature of the documents being photographed and is approximately 2500 pages per shift.

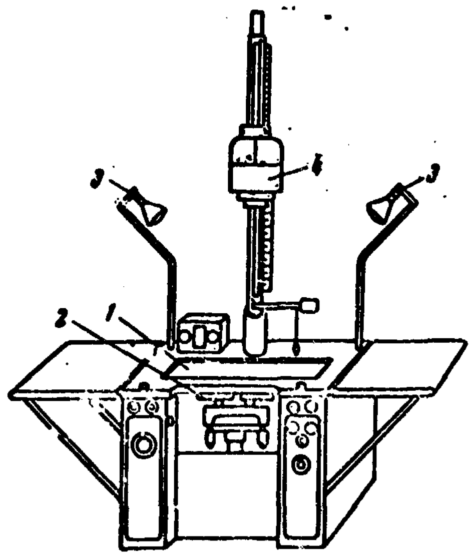


Fig. 1. UDM-2 microfilming setup. 1) Screen; 2) small table for books; 3) illuminating lamps; 4) camera.

Among other microfilming units of the static type we might name the Dokumator (models IV and V) produced in GDR by the nationalized firm Karl Zeiss. This unit is designed to use 35 mm roll movie film. The maximal reduction ratio is 26; the maximal format of the document being photographed is 814 × 1152 mm.

The Soviet microfilm unit RUST-3 is produced for microphotocopying of drawings and other documents of large format (to 814 × 1152 mm). The photographing is accomplished on roll unperforated photographic film of

70 or 35 mm width. The maximal reduction ratio is 12.8.

The microfilming equipment for dynamic filming differ externally from the microfilming units for static filming. They have no screen, no external illumination, they are considerably smaller in size.

The principle of dynamic, or slit, filming consists in the illumination and photographing at each moment of a narrow strip (slit) on the original, which moves continuously with various segments of the surface falling in the illuminated field. The light reflected from the original passes through the lens onto the light-sensitive layer of photographic material, which moves synchronously with the original; the photographic film moves slower by a factor equal to the reduction factor.

The equipment for dynamic filming, having smaller dimensions and requiring considerably less electrical power, provide a productivity several times that of the static filming units. The dynamic units may be used to photograph originals in the form of individual sheets and the length of the photographed original is unlimited.

The Moscow Design Bureau for Cine Equipment, MKBK, has developed three types of dynamic filming equipment. The NM-1 unit operates as follows. The stack of documents is placed on the self-feeder, which feeds them one after another over a transporter to the filming slit. The rate of movement of the documents (apparatus productivity) is 0.6 m/sec, i.e., in one hour up to 9000 sheets of length 200 mm may be photographed. The maximal width of the photographed document is 300 mm, the length is unlimited.

Text located on both sides of a sheet may be photographed in a single pass of the document. The photography is accomplished on a single or simultaneously on two roll photographic films of width up to 16 mm. The reduction ratio is 24, 32, and 43.

The NM-2 microfilming unit (Fig. 2) is intended for photographing

on roll photographic film of width 35 or 70 mm engineering drawing originals and other large-format documents. The width of the photographed original may be up to 850 mm. Filming may be accomplished in reflected or transmitted (for transparent originals) illumination. The reduction ratio using photographic film of 70 mm width is 10 or 13, for 35 mm film the reduction ratios are 19 or 26. The rate of travel of the document through the photographic slit (unit productivity) is 0.3 m/sec.

GRAPHIC NOT
REPRODUCIBLE

Fig. 2. NM-2 microfilming unit.

Figure 3 shows the NM-3 portable microfilming equipment. It occupies an area of 270 × 410 mm and has a height of 330 mm. The width of the photographed documents (in the form of individual sheets or small notepads) is up to 300 mm, the length is unlimited. The unit is equipped for photographing in reflected or transmitted (tracing paper) light. The document rate of travel is 0.3 m/sec. The photographic film width is 16 or 35 mm; the reduction ratio is 9 and 20.

Still another unit for dynamic microfilming, AShM-1300, has appeared recently, intended for microphotocopying of documents of very large format – width to 1300 mm with unlimited length.

After filming the photographic film is subjected to chemical proc-

essing in solutions, washing, and drying. The Mikrat photographic films are processed in ordinary concentrated developing solutions for black and white photographic materials. The developing time is 4-8 min; after developing the film is rinsed in water or immediately placed in a fixing solution and later washed. Very careful washing is required in order to provide storage for several years, in order to practically completely remove the remains of the fixing salt. Washing of the photographic film continues for about 10 minutes in tap water. After this the film is dried in room conditions or in a drying box using slightly heated air.

GRAPHIC NOT
REPRODUCIBLE

Fig. 3. NM-3 microfilming equipment.

The chemical processing of the roll photographic film is performed in special devices or developing machines which are adapted for providing uniform contact of the entire surface of the light-sensitive layer with the processing solutions and the wash water. Devices with spiral winding of the photographic film may be used for this purpose; the spiral coil with the photographic film is submerged in a tank with the solution, where the processing solutions enter the gaps between the spiral turns. The Karacharov Plastics Plant produces such a tank, designed for processing 30-meter lengths of 35- or 16-mm photographic film.

The manual developing devices are advisable for use when it is necessary to process small quantities of photographic film. For large quantities (more than 300 m per working day) special developing machines are used. In the latter the roll photographic film passes continuously through all the stages of the chemical processing: development, fixing, washing, and drying. The developed and dried photographic film is gradually wound into a roll. The Leningrad Cine Equipment Plant Leninkap produces the 60P-4 developer intended for chemical processing and drying of 16 and 35 mm perforated or unperforated Mikrat photographic film. The productivity of this machine (rate of photographic film travel) is 23, 47, 93 or 185 m/hr. The productivity is determined by the necessary duration of stay of the photographic film in the developing solutions. For photographic film of the Mikrat type the developing lasts 4-8 min, which corresponds to film travel rates through the 60P-4 developer of 47-93 m/hr. After the chemical processing the microfilms are subjected to careful visual inspection in order to eliminate possible mechanical damage (scratches) and other defects which can influence the readability of the reproduced documents.

Any desired number of roll positive copies may be printed by the contact method (without changing the image scales) from the resulting roll negatives of the document microphotocopies. Special high-production copying devices are used for this purpose.

Soviet industry produces three types of copiers for printing microphotocopies on roll photographic film: MKP-1, MKP-2, and MKP-3. The MKP-1 and MKP-2 copiers differ in that the first is intended for printing on roll photographic film of width 16 and 35 mm, while the second is for 35 and 70 mm.

Figure 4 shows a general view of the MKP-1 copier. The printing rate on this unit is 150 m/hr. The MKP-3 copier is a very portable de-

vice (Fig. 5), intended for printing on 16 and 35-mm photographic film. The printing rate is 50 m/hr. The AKD-55 mass-produced copier (Fig. 6) may be used for printing individual frames with a microimage. It is recommended for selective printing of individual frames of dimension 24 × 36 and 18 × 24 mm from 35-millimeter roll negative onto 35-millimeter positive roll photographic film. The unit is very portable, has dimensions of 235 × 195 × 172 mm; a darkroom is not required for operation.

Soviet industry produces photographic films (see above) and all the necessary equipment for large-scale preparation of microphotocopies on roll photographic film with reductions to 20 times and more. Microphotocopies on roll photographic material, the so-called microfilms, have found wide application in this country. The technological process for their production is usually termed microfilming. In many cases this process is based on the highly developed cinematographic experience; this obviously explains the wide usage of microphotocopying in the roll version. However, the arrangement of the microphotocopies on rolls has several essential drawbacks, which complicate the use of the microphotocopies both for manual operation and for creating automatic systems for information retrieval. It is difficult to work with the strip in roll form, special devices and time are required for loading the roll into the viewing device (for reading). If the required material is located at the end of the roll it is necessary to reroll the reel, which is time consuming and requires the use of specialized equipment.

It is impossible to design a flexible, efficiently expandable system for information retrieval and output (it is impossible to introduce new microphotocopies without cutting the roll) on the basis of the roll arrangement of the microphotocopies.

These drawbacks resulting from the arrangement of the microphoto-

GRAPHIC NOT REPRODUCIBLE

Fig. 4. MKP-1 copier.

GRAPHIC NOT REPRODUCIBLE

Fig. 5. MKP-3 copier.

GRAPHIC NOT REPRODUCIBLE

Fig. 6. AKD-55 copier.

copies on a roll are completely eliminated with arrangement of the microphotocopies on individual segments of photographic material — microcards.

Thus, the reduced photographic copies of documents — microphotocopies are divided into two forms: microfilms and microcards.

3. Equipment and Technology for Mass Production of Microcards

Microcards have not yet found wide practical application in the Soviet Union, therefore the name itself is somewhat unusual. By a microcard we mean a card fabricated from photographic material or printed by a polygraphic method, whose surface has the image of microphotocopies of documents. The microcard may be fabricated on a transparent (transparent film) or nontransparent base (white paper); in the latter case it is possible to arrange the microphotocopies on both sides of the card.

Outside the Soviet Union it is usual to use the term *microcard* for microcopies arranged on cards made from photographic paper (on a nontransparent base). The exact same cards, made from photographic film (on a transparent base), are termed *microfiche*, while those fabricated from paper with reduced images of the documents printed by a polygraphic method are termed *microprints* [5].

In our translated literature we encounter the term *diamicrocard*, i.e., microcards on a transparent base which may be read in transmitted light, and *epimicrocards*, microcards on a nontransparent base which are read in reflected light.

The simplest is the microcard with a single row arrangement of the document microcopies. Such microcards may be fabricated by cutting the roll microfilm into individual pieces corresponding to the length of the microcard. Considerably more complex is the preparation of micro-

cards with arrangement of the document pages in several rows; with this arrangement the microcard takes on the form of the conventional card catalog card. The number of pages (frames) of the documents and rows on the microcard depends on the size of the originals, the reduction ratio, and the microcard format.

As a rule, the microcopies of documents must be arranged on the microcards just as the words and lines of a manuscript or printed text are arranged, from left to right and from top to bottom.

The most widely used microcard dimension is 75 × 125 mm (Fig. 7) – the format of the standard bibliographic card. Less widely used is the 105 × 148 mm format. In the systems for mechanized information search which use as the information carrier microphotocopies of documents, quite different formats are permitted, depending on the conditions of system operation.

In principle, the same photographic materials may be used for preparing the microcards as were used for preparing roll microphotocopies with the single essential difference that the photographic film must not be curled, i.e., it always remains flat, regardless of temperature oscillations and humidity changes of the surrounding air. Considering that the photographic film is a system of several layers with different physical and mechanical properties, the satisfaction of this requirement is related with increased complication of the fabrication process for the photographic film itself. At the present time the back side of the photographic film opposite the light-sensitive layer is coated with the same thickness of an anticurl layer, which leads to a general thickening of the material. Development is continuing of a special lacquer coating for the photographic film to ensure its flatness.

Three technological processes are known for fabricating microcards with multi-row arrangement of the document microcopies.

GRAPHIC NOT REPRODUCIBLE

Fig. 7. Microcard (multiple-frame microcard).

1. The roll negative of the document microphotocopies is cut into individual strips corresponding to the length of the rows on the microcard. Then these strips are arranged on a glass as they are to be positioned on the microcard copy and from this montage contact prints are made of the required number of microcards. This method is quite acceptable if we are not required to retain the negative, which must be cut into individual pieces. This method essentially does not require any special equipment and is available to any photographic laboratory which has equipment for contact printing. The montage of the microcards from the negative film strips is a very simple manual operation.

2. The microfilming of the documents is performed with multi-row arrangement of the documents. For this we use a special wide photographic film corresponding to the width of the microcards. After exposure the film is processed by the usual method, and the required number of microcards are printed from the resulting negatives by the contact method. Thus, the multi-row arrangement of the microphotocopies of the

documents is accomplished here in the photographing process. For this purpose we require specialized microfilming equipment for static filming with stepped transverse movement of the photographic film for photographing each subsequent row on the microcard. Such equipment has comparatively low productivity, particularly in comparison with the microfilming equipment for dynamic filming. For mass production of microcards using this method it would be necessary to build specialized microfilming equipment, since the equipment which we produce with single-row arrangement of the microcopies of the documents cannot be utilized.

3. This method is of most interest, since it is based on the existing high-productivity microfilming equipment with single-row arrangement of the document microphotocopies. While in the second method the multi-row arrangement of the document microphotocopies was accomplished during the negative process (during microfilming), here, during the positive process a contact print is made from the roll negative of the microcopies onto a wide positive photographic film (corresponding to the width or length of the microcards) so that in each row of the microcard there is printed the corresponding strip (piece) of the roll negative.

Thus, in place of the equipment for stepped microfilming here we require equipment for stepped multi-row contact printing, which is simpler than the equipment for stepped filming.

It must be emphasized that in this method we use the roll negative prepared on the conventional, widely used microfilming equipment. From the roll negative we can print microcards or roll positive copies - microfilms.

A prototype of a manual copier for printing microcards from roll microfilms has been built at VINITI (Fig. 8). The equipment permits printing microcards on roll photographic film or photographic paper of

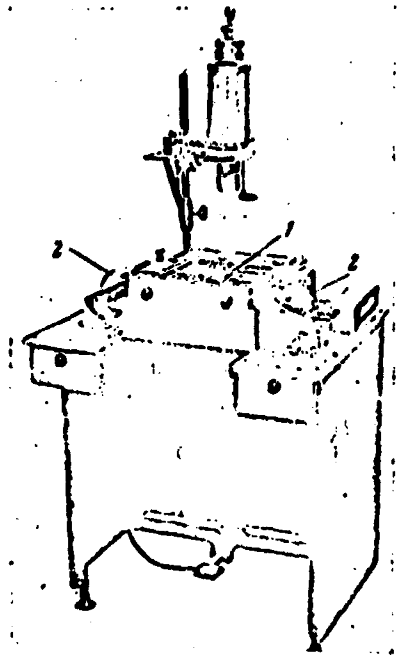
75-mm width. Printing is performed from roll negatives onto perforated or unperforated photographic film of 16 or 35-mm width.

The positive photographic film passes through the contact table in the chamber 1. The negative photographic film is wound on the reel 2, passes under the glass which presses the two films together at the moment of printing. After printing each successive row the chamber is shifted by the width of one row and the negative photographic film is shifted for printing the succeeding microphotocopies. The prototype of the manual copier enables the printing of the order of 100 microcards per working day.

A prototype of a copier for printing microcards from roll microfilms has also been fabricated at NIKFI. The equipment is designed for printing from 35-mm perforated photographic film onto roll photographic material (film or paper). The Moscow Design Bureau for Cine Equipment, MKBK, following the specifications of VINITI and NIKFI has developed a semiautomatic copier for printing microcards from roll microfilms (Fig. 9). The productivity of the unit is of the order of 200 four-row microcards per hour. The entire row being printed is viewed on the screen 1 with a 5-fold magnification, or individual microcopies are viewed with a 25-fold magnification. The equipment provides for selective printing, i.e., printing individual frames from the negative.

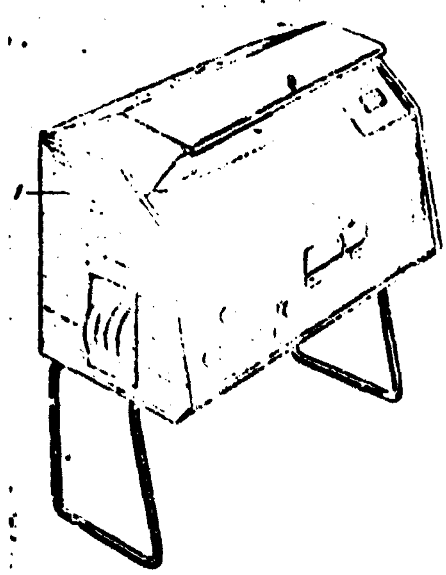
The chemical processing of wide photographic film with microcards printed from the roll negatives is accomplished just as the usual roll photofilm with positive microphotocopies. Equipment is being developed at VINITI for developing and drying roll photographic films with microcards and for cutting the roll photographic film into individual microcards.

Very promising for the efficient and economical production of copies from microcards is the use of diazo film, on which copies may be



GRAPHIC NOT REPRODUCIBLE

Fig. 8. Prototype of copier for printing microcards from roll microfilms.



GRAPHIC NOT REPRODUCIBLE

Fig. 9. The AKPM (developed by MKBK) copier for printing microcards from roll microfilms.

made from microcards directly at the place of storage immediately after selecting the required microcard.

In addition to the microcards on photographic material which permits producing microcopies with large magnifications, of interest are the materials printed by the typographic method with some reduction of the text dimensions – up to 5 times.

In addition to the size reduction, such microeditions are significantly cheaper than the conventional typographic editions (the cost is reduced approximately in proportion to the square of the reduction ratio). Special optical devices are required for reading the reduced text. Microeditions are particularly advantageous for long runs.

III. MECHANIZATION AND AUTOMATION OF INFORMATION SEARCH

It is difficult to imagine a system for mechanizing information search with the traditional storage of paper documents on shelves and racks, when they are dispersed in space. It is a different matter if huge masses of documents are compressed into a small volume and the required document is automatically selected from an array, made available for reading, or a copy is made from it. Microphotocopying has created the primary basis for the mechanization and automation of information search – the reduction of information into a comparatively small volume. The mechanization and automation system based on document microphotocopies usually contains two elements: document microphotocopies on a roll (on a tape) with some sequential arrangement of the materials, or in individual segments of photographic material (microcards), and the search labels for the corresponding documents.

The search labels may characterize the document content with varying degree of detail. In the simple cases these are brief bibliographic summaries, while in the complex cases they are a detailed description

of the document which provide for accomplishing a logical search. The search labels are arranged alongside the document microphotocopies or are located separately from them.

The first case is advantageous when the selection is made from a large number of documents with respect to a comparatively small number of labels. The possibility is created for performing a single-step search — the desired document is selected simultaneously with the search labels.

With a large number of search labels for a comparatively small volume of documents it is advisable that the search labels be separate from the documents. Then the search is accomplished in two stages: first we find the storage place of the documents (address) from the search labels, after which the document itself is selected.

Three methods of writing the search labels are known: perforation, photographic, and magnetic. The perforation writing of the search labels is accomplished by punching in a definite combination openings over the entire field or along the edges of paper cards or tape. The document microcopies are mounted in a window in the paper cards; such combined punch cards with microcards are usually termed aperture punch cards. The photographic writing of the search labels is a combination of dark and transparent segments on photographic material. Very high code density can be achieved by this method (see, for example, Fig. 18). The magnetic recording of the search labels is accomplished in a magnetic layer by combining magnetized and nonmagnetized segments of a tape or card. The magnetic layer may be arranged alongside the photographic strip where the document text is recorded. A special photomagnetic material on which magnetic tracks are located beside the photographic layer is being developed at NIKFI.

In principle it is possible to combine the photographic and mag-

netic recording of the code, of which the first carries invariable (constant) labels, and the second carries variable labels.

Search systems are known which use roll microphotocopies arranged on individual segments of photographic film (microcards). The roll arrangement has the following significant drawbacks: a) the impossibility of adding or subtracting from the roll any portion of the material (without cutting the roll and subsequent splicing), i.e., the impossibility of obtaining a flexible, adaptive system; b) in order to accomplish sequential search with the soft material located at the end of the roll, it is necessary to rewind the entire roll; c) the search rate (rate of movement of the photographic film tape) is limited by the fact that during rewinding of the roll there are two variable masses – one roll diminishes (from which the film is unwound), the other grows.

These drawbacks are completely eliminated when using microcards, which are easily added to as new material arrives, unnecessary microcards are taken out, during search there is no need for looking sequentially through the entire array of material; since the individual microcards have very little weight, inertialess systems may be created for their transportation, providing for adequate speed of operation.

However, the roll version has certain advantages: the possibility of losing individual document pages is eliminated; the roll is more easily subjected to mechanized sterilization processing which is required for long-term storage of the microphotocopies.

The systems for information storage and search based on microphotocopies of the documents differ in the method of recording the code of the search labels (perforation, photographic, and magnetic); arrangement of the search label code separate from the document microphotocopies (two-step search) or together with the document microphotocopies (single-step search); arrangement of the microphotocopies on roll pho-

tographic material or on individual segments of the photographic material - microcards.

1. Manual Search

Roll negative and roll positive. The microphotocopies of the documents are arranged on roll photographic film of 16, 35, or 70 mm width. Most widely used is the roll film of 35-mm width. The negative provides for long-term storage of documents, and from the negative there are printed out positive copies which serve for reading the documents periodically as they are needed. The roll length is determined by the nature of the documents and the frequency of referral to them. In the case of low loading of the information system (archive materials) the document microcopies may be grouped in large rolls of up to 300 meters - roughly more than 15,000 pages with a reduction ratio of 15.

The most common roll lengths are 30 meters for the negatives and 3-4 meters for the positives. The rolls with positive microphotocopies are stored in wooden cabinets with sliding drawers (Fig. 10) which are equipped with cells in the form of a honeycomb. Metal or plastic boxes are also used for storing the rolls. The positive copies may be stored in polyethylene bags or without any packaging if dust is kept away from them. The rolls are numbered, arranged in proper order, and indexed using a card file.

This system for microphotocopy storage and retrieval has been installed in the Lenin Library, the Main Archive Administration, Gidrometsluzhba and other installations.

Roll negative and positive on single-row microcards. The roll negative remains in the form described in the preceding section, while the roll photographic film with positive microphotocopies of the documents is cut into individual segments - microcards of established length. The

microcard permits the inclusion of several microcopies of documents on the same theme and is always equipped with a heading frame (Fig. 11) including, for example, a microcopy of the title page of a journal, a code or working number which is used to code the article. The latter are made large enough so that they are visible on the film without magnification. The microcards are stored by stacks and are collected into card files.

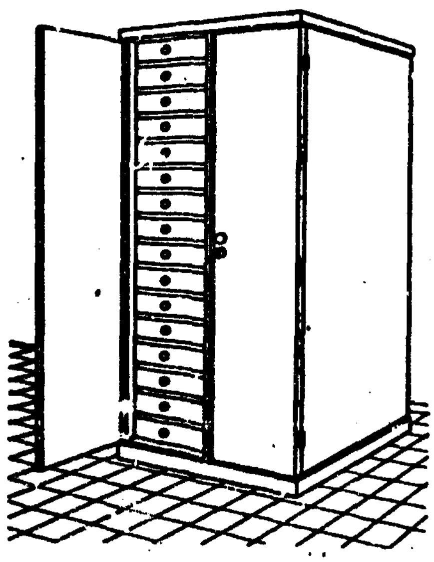


Fig. 10. Cabinet for storing microfilms.

In the GDR (Dokumator system) albums are used on the sheets of which pockets are made for the microcards.

In comparison with the roll systems the microcard systems have the following advantages: a) it is more convenient to find all the required material in the card file; b) it is easier to load the microcards into the equipment for reading the microcopies; c) only the required document is located on the microcard. If a single article is located on each microcar[illegible]tack of microcards with the annual set of articles of a journal will have a thickness of the order of 50 mm (assuming that each number of the journal has 25 articles).

Negative and positive on single-row microcards. In contrast with the system described in the preceding section, in this case the roll film with the negative microphotocopies is also cut into individual pieces. This makes it easier to select materials for printing the posi tives; it is difficult to accomplish selective contact printing from the roll negative (on the existing copying equipment). At the same time the location of the negative microphotocopies on the segments increases considerably the possibility of their loss.

Roll negative and positive on microcards with multi-row arrangement of document microcopies. In contrast with the simple microcards with a single-row arrangement of the microphotocopies, here the latter are arranged on the microcards themselves in several rows from left to right and from top to bottom. This arrangement leads to a convenient format of the microcopy, similar to the widely used format for catalog cards. An example of the multi-row microcards printed from roll negatives are the so-called summary microcards with the image of a large number of microcopies of the catalog cards or other card-catalog materials (see Fig. 7).

With a reduction ratio of 14, on a single summary card of 75 × 125 mm format [6] there are located 144 standard bibliographic cards, with a reduction of 25 there are 350 such cards, and with a reduction of 33 there are 700 cards. The summary microcards, retaining the usual order of sequential arrangement of the cards in the card file, significantly reduce its volume and facilitate operations. Even with a 14-fold reduction, one million bibliographic cards, transferred to the summary microcards, fit into 5 card-catalog drawers (1500 cards in one drawer); with the usual arrangement this same volume would occupy 650 drawers. The introduction of the summary microcards also solves the problem of coding and duplication of the cards.

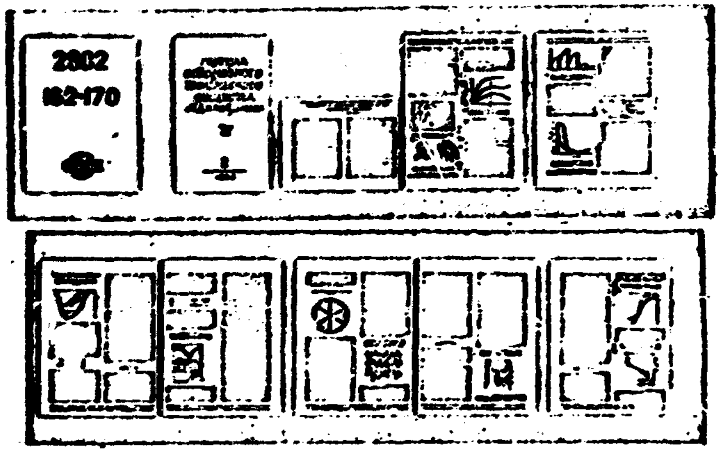
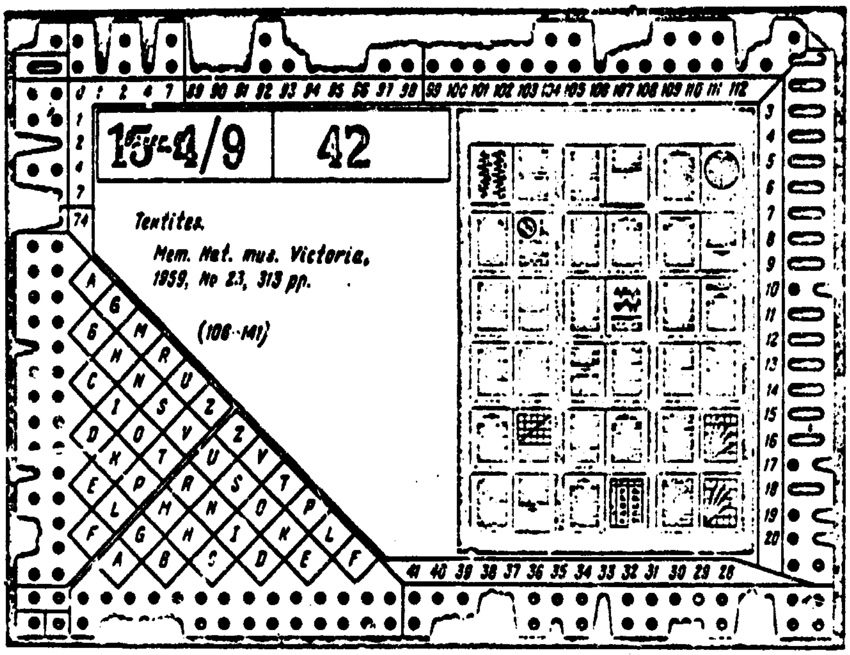


Fig. 11. Single-row microcard.

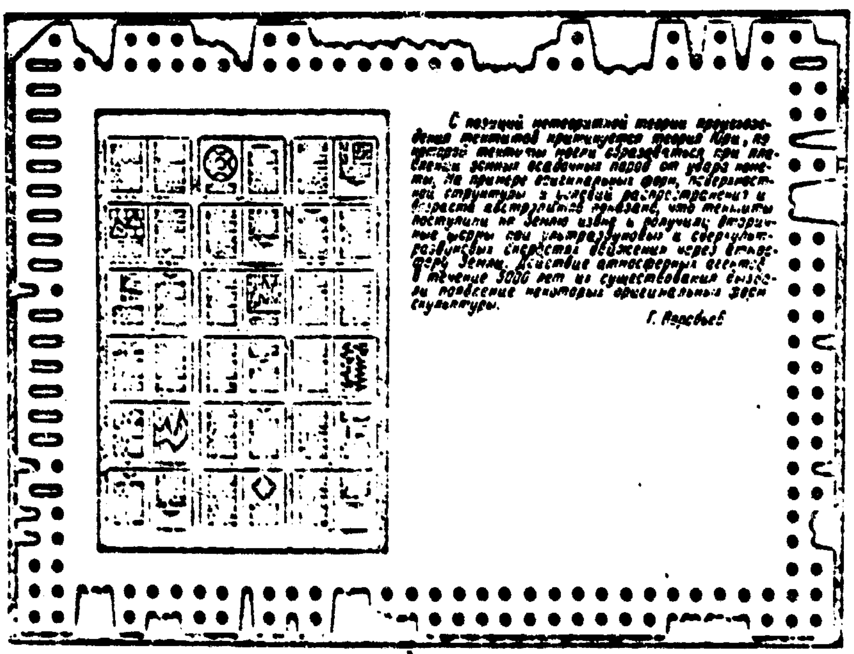
Using the scheme described for storage and retrieval – negative on a reel, positive on a microcard – the collection of any documental materials may be conveniently organized so that materials on the same subject will be located on an individual microcard. The practical implementation of this scheme requires a copier for printing microcards from roll microfilms. The sample summary microcard shown in Fig. 7 was produced on a prototype copier developed at VINITI.

Aperture punchcards are a combination of punch cards and microcards [7]. A window in which the microcard is mounted is made in the punch card. Three window dimensions are provided for punch cards with edge perforations of 207 × 148 mm format: 35 × 48, 35 × 105, and 75 × × 105 mm, which correspond to 4, 12, and 40 pages of journal text for a 15-20-fold reduction. Mounting of the microcards is accomplished by applying a narrow strip of adhesive tape of 5-6 mm width along the perimeter of the window so that a portion of this strip is on the surface of the punch card while the other part is on the surface of the microcard. The microcard slips into the window and lies in the same plane with the punchcard. The adhesive tape is produced on thin (less than 0.1 mm) transparent film, one side of which is coated with the adhesive layer and the tape is wound into a roll with the adhesive layer inside. The film base is cellophane (produced by the Mosgorsovnarkhoz firm) or polyethylene (Okhtinsk Chemical Combine, Leningrad). Adhesive tape on a very thin (0.05 mm) and strong lavsan film is being readied for production.

The aperture punchcard (Fig. 12a) carries a bibliography, search labels which are entered by cutting out edge openings, microcopies of the documents, and an abstract (Fig. 12b). Operation with a card file of manual aperture punch cards is accomplished as follows: on the basis of the posed questions a search is made for the necessary cards using a



a



b

Fig. 12. Aperture punchcard. a) Face side; b) reverse side. "A critique of the Urey theory, according to which tektites might have formed as a result of melting of terrestrial sedimentary rocks on collision with a comet, based on a nonmeteoritic theory of the origin of tektites. With reference to an example of original shape, surface structure and conditions of propagation and growth of australites, it is suggested that tektites came to the earth from elsewhere and acquired secondary forms at supersonic and hypersonic velocities through the atmosphere. Action of atmospheric agents, over the 5000 years of their existence, gave rise to the appearance of certain original sculptural forms."

— G. Vorob'yev.

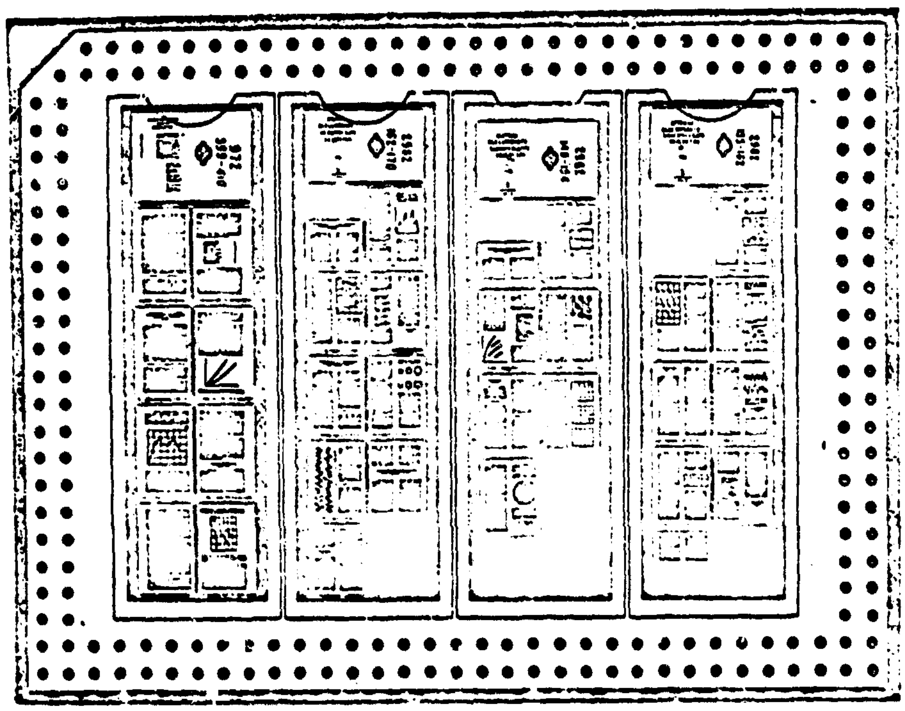


Fig. 13. Klyassernyy punchcard.

selector needle (the cards are arranged at random in the card file), the abstract is read on the selected cards, and if necessary the card is placed in a conventional reader for reading or copying (see above).

Klyassernyy punchcards. Small pockets are provided on one side of the punchcard in which microcopies of the documents are placed in the form of microcards. While the aperture punchcards do not provide the possibility of introducing additional information or removing obsolete materials, the klyassernyy cards have a clear advantage in this respect. The pockets are made from a transparent thin film, which permits examining the heading on the microcard or reading the entire text on a reader using reflected light without taking the card from the pocket. The klyassernyy punchcards may be useful in office work and in all other cases related with the movement of documents concerning a single common question.

Figure 13 shows a photograph of a klyassernyy punchcard from the side of the pockets for inserting the microcopies. The pockets of the klyassernyy punchcards provide reliable protection of the microcard surface from mechanical damage.

2. Some Systems for Highly Mechanized Automated Search, Based on the Use of Microphotocopies of Documents

The principle of highly mechanized search is that using a coded form of the search labels a servo device selects the document microcopy (if it is located alongside the search label code) or indicates its storage location (address) in order to accomplish the search the material (photographic film, magnetic tape, magnetic card) on which the search label code is recorded is transported relative to the servo unit. The latter is a block of photodiodes or a block of magnetic heads with corresponding electronic logic circuit for receiving and transmit-

ting signals. To accomplish the search a request is entered into the searching unit.

As indicated above, the information search systems are based on tape (roll) or cards.

Search systems using microcopies on roll photographic film. The Rapid Selector search system [8] using roll photographic film began operation back in the 30's at MIT (USA). The document microcopies and a photographic code record of their search labels are arranged on 35-mm film. The film passes through a servo unit with a speed of 100 m/min (3000 pages of documents). The photographic code of the search labels (up to 240 cells) is located alongside each document. When the required document is detected its photocopy is automatically printed without slowing the speed of the photographic film.

More advanced models of the Rapid Selector system are presently being developed with higher film travel speed (by a factor of 2-4).

The Flip machine of the Benson-Lehner firm (USA). Roll 16-mm perforated photo film is used. Alongside the microcopies of the documents the 32-symbol photographic code is recorded. On a roll of length 376 meters there are 72,000 frames, the roll speed is 1.5 m/sec. The selected frame is projected on a screen with a magnification of the order of 50.

The File search machine of the FMA firm (USA). On roll photographic film of length 330 meters there are 32,000 pages and scanning is accomplished with a speed of 6400 pages/min (photographic film speed is 1 m/sec). Up to 392 code symbols may be recorded on each frame.

The Poisk I and Poisk OK search units using 35-mm unperforated photographic film are being developed at VINITI of the Academy of Sciences USSR.

The Filmsort system of the Minnesota Mining and Manufacturing firm

(USA) is based on a combination of machine punchcards of 187 × 82 mm format and microphotocopies of the documents [9]. The latter are arranged on a microcard of size 35 × 48 mm mounted in the window of the punchcard (Fig. 14).

More than 50 columns on the 80-column punchcard may be used for recording the search label code. Standard punchcards and an assortment of counting-perforating machines are used. However, in order to avoid rapid wear of the aperture cards it is better to perform the search using conventional punchcards in order to then select manually the required aperture punchcards. The Filmsort system has found application in construction drawing documentation services.

A manual machine with an average productivity of 400 cards per hour (Fig. 15) and an automatic machine (Fig. 16) with an output of 2000 cards per hour are used for mounting the microcards in the punchcard windows.

Contact copying of the microphotocopies onto the aperture punchcards is accomplished on diazo film, mounted in the punchcard. The aperture cards with unexposed photographic film (diazo film) are delivered in finished form; to obtain copies from the aperture punchcards use is made of a special portable copier for printing from card to card (Fig. 17).

The Minicard system of the Eastman Kodak firm (USA). The microphotocopy of the document or group of documents is arranged together with the photographic code on the microcards [10] (Fig. 18). The microcard size is 16 × 32 mm. The microcards contain up to 12 document pages with a reduction ratio of 60 and 2730 symbols of the search label code. There is available a negative Minicard microcard from which any desired number of positive copies to be used in the search system may be printed. All the microcards on a common subject are threaded on a common rod

(there is an opening in the upper part of the microcards for this purpose) and are stored in this form.

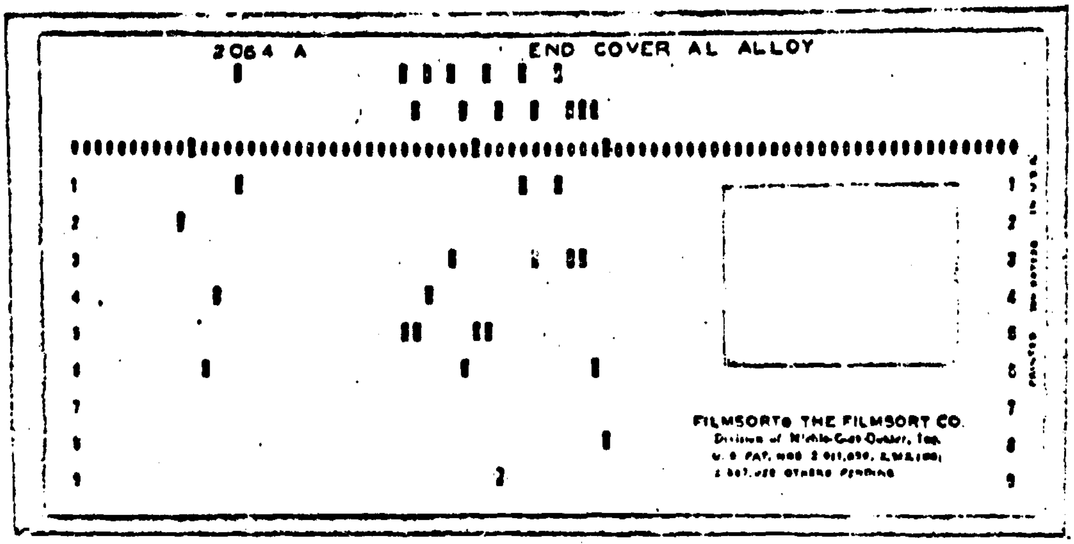


Fig. 14. Aperture punchcard of the Filmsort system.

GRAPHIC NOT REPRODUCIBLE

Fig. 15. Manual machine for mounting Filmsort aperture punchcards.

GRAPHIC NOT REPRODUCIBLE

Fig. 16. Automatic machine for mounting Filmsort aperture punchcards.

Such an arrangement provides a very flexible possibility for adding to or removing part of the material. To accomplish the search the rod with microcards is introduced into a search unit with a handling capacity of 1800 units/min. Copies are automatically printed from the

selected microcards, or the text is projected on the screen of a reader. The system provides extremely compact document storage.

GRAPHIC NOT
REPRODUCIBLE

Fig. 17. Unit for printing from card to card.

The Minicard system was first introduced into operation in 1958 in the USA Department of the Air Force in the Pentagon. Later several governmental departments of the USA, including the Central Intelligence Agency, converted to this system. The entire complex consists of 13 individual photographic, optical, electronic, and mechanical units of a total cost from 2.5 to 3.5 million dollars.

The Magnavue system of the Magnavox firm (USA). Here the search label code is recorded magnetically [11]. This provides an advantage relative to the photographic method of recording in that the possibility is provided for easily introducing corrections and additions.

GRAPHIC NOT
REPRODUCIBLE

Fig. 18. Microcard of the Minicard system.

The microphotocopies of the documents and the magnetic matrix are arranged on 25 × 75 mm size microcards. The card base is lavsan which has considerably higher strength in comparison with the widely used triacetate base. The document microcopies occupy one-third of the card; the remaining two-thirds are covered with a magnetic layer with 3000 matrix cells. The cards pass through a scanner with a speed of 5400 units/min. From the magazine where they are stored they are taken (sucked) by a rotating drum to the surface of which a vacuum is applied. This method of transporting minimizes the card wear and provides a very high search speed (the peripheral speed of the drum rotation is on the order of 10 m/sec), which is impossible with the use of roll photographic film.

The cards selected during the scanning are sorted, the microcopies are read on the screen of a reader or magnified copies corresponding to the dimensions of the original documents are printed from them. The text image may also be transmitted over a television channel.

The Microcite system at the National Bureau of Standards (USA) is based on a combination of transparent (superpositioning) punchcards and microphotocopies of documents. The search symbols for 18,000 documents are stored on manual transparent punchcards, the microphotocopies of these documents are arranged on large sheets of photographic film of 360 × 360 mm format. The sheet of photographic film with the microphotocopies is placed on a drum (Fig. 19) which is interlocked with coordinate bars indicating the space corresponding to the number of the selected document [12, 13].

The search is accomplished by manual selection of the superposition cards corresponding to the interrogation. The cards are aligned on top of one another, and the coordinate bars are adjusted to the lighted segments on the upper card. Depending on the position of the bars (co-

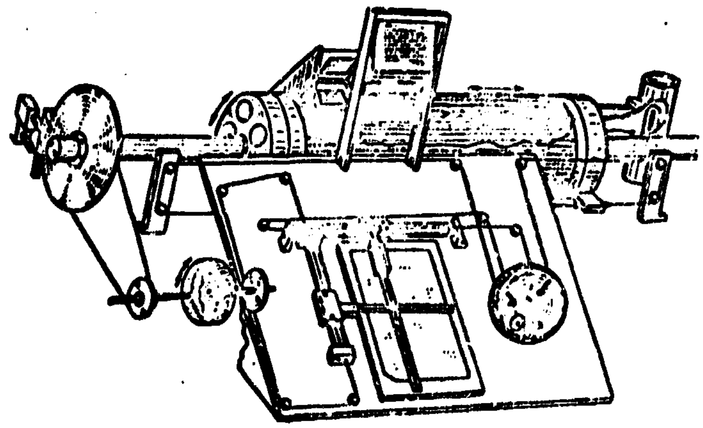


Fig. 19. Schematic of the Microcite system.

ordinates of the light gap) the drum and the photographic film mounted on it take a position where the microcopy of the desired document is in the framing window and its magnified image immediately appears on the screen.

The Verac system of the AVCO firm (USA). The microphotocopies of the documents are arranged on cards made from photographic film of 200 × 200 mm format. With a reduction factor of 140 there are 10,000 microcopies on a single card; 100 cards contain 1 million documents [1, 8]. To obtain microphotocopies with such large reduction factors use is made of the photochromic process. There are no recordings of the search labels on the cards. The search for the required documents is accomplished with the aid of the manual finder (for example, using superposition cards) or a computer. The search amounts to determining the coordinates of the microcopy of the document on the card, after which the required document is found within 2 seconds. The system permits obtaining microcopies of the documents on photographic film segments of width 35 mm or magnified copies, right up to the full-size dimensions of the document. The microcopy may also be examined using a television channel.

With regard to compactness of document arrangement the Verac system surpasses all the previously mentioned systems. In addition, having very high capacity, it can provide rapid search of a large array of documents with respect to a large number of labels without recourse to complex electronic machines. With document reduction factors greater than 100, when thousands of documents are arranged on the small area of the photographic film, there is no need for complex devices for transporting during search, which simplifies considerably the search device itself. Thus, the high reduction ratio of the documents using microphotocopying has solved another important problem in addition to the problem of compact document storage – the problem of creating simple high-speed systems.

The Walnut system of the IBM firm (USA) solves the problem of transporting photographic film with microphotocopies arranged on the film so that the wear of the photographic film is reduced to a minimum. Protection of the photographic surface from mechanical damage during its passage through the scanner in the searcher is a very complex problem which is solved in this system as follows: a) the search label code is separated from the document microcopies and is arranged separately on magnetic tape; b) the microcards with the microphotocopies of the documents are not transported – they are stored permanently in special magazines and are taken out (at low speed) for copying the information [14].

The microphotocopies are reduced in two steps. First microfilm on 35-mm roll photographic film is prepared, which is then again reduced during printing of the image on the microcards; the over-all reduction ratio of the document is 35. On each microcard there are arranged 3 rows with [illegible] microcopies per row – 99 pages in all. The microcards are arranged vertically in special plastic magazines holding 50 units. The

magazine dimensions are 26.7 × 39 mm, height 397 mm (containing 4950 microcopies). The magazines are mounted on a drum in groups of 200, which thus provides for a minimal information capacity of the system of 990,000 document microcopies. At most the system can contain more than 100 such card-file drums. The microcards never leave the magazine, therefore their wear is eliminated. The search label code is recorded separately on magnetic tape. The search for the required documents is accomplished in two stages: first on the magnetic tape the location of the document is established, after which the drum begins to rotate and the magazine where the microcard with the given document is located is fed to a special set of fingers for its removal. After the microcard is removed from the magazine, the required microcopies are illuminated by ultraviolet rays and the image is printed on an aperture card carrying unexposed photographic film of the Kalvar type. Immediately after exposure the image is developed by heating, and the information is delivered on the aperture punchcard. The microcopy may be read with the aid of a reader or an optical magnified copy may be printed from it. Only 5 seconds is required to obtain such a copy from any interrogated document. On the average 600 aperture cards with four microcopies on each of them may be output in 1 hour.

IV. READING AND EFFICIENT CODING OF MICROPHOTOCOPIES

Reading of the microphotocopies is accomplished on special readers intended for magnifying them to full size. Stationary, table-top, portable, and even pocket readers are available. With respect to nature of the illumination we differentiate two types of equipment: for reading in transmitted light (microcopies on a transparent base) and in reflected light (microcopies on an opaque base). There are available also combined readers for reading in transmitted and reflected light.

The type 5P01 Microfot reader (Fig. 20). This table-top type equipment is produced in serial quantities by the Moscow Cine and Electromechanical Plant - KEMZ. It is intended for reading in transmitted light of roll microfilms, microstrips (segments of 35-mm photographic film). A special attachment is fabricated at VINITI for reading microcards and aperture cards (Fig. 21). The screen size is 300 × 360 mm. The device is equipped with two Jupiter-8 and Jupiter-12 lenses for reading with magnifications of 10 and 16. The microcards and microstrips are fed into the equipment for reading much more simply than the microcopies on roll photographic film.

The AChS-1 reader (Fig. 22). An experimental model has been developed and fabricated by the Moscow Design Bureau for Cine Equipment and serial production is planned. The equipment is stationary, intended for reading microphotocopies on roll photographic film of 35 and 70-mm width or microcards. The following magnifications are provided: for 70-mm photographic film - 7.7 and 13; for 35-mm photographic film - 18 and 30. The screen size is 850 × 650 mm.

The AChM-2 reader. An experimental model has been developed and fabricated by the Moscow Design Bureau for Cine Equipment, preparations are being made for serial production. The portable model (Fig. 23) is intended for reading microphotocopies on roll photographic film of 16 and 35-mm width or microcards. The screen size is 550 × 550 mm, the image is projected on the table. Magnifications of 15 and 26 are provided. This equipment may be used not only in institutions, but also under field and home conditions. It may also be used as a projector or photographic enlarger.

The AChK-3 pocket reader. An experimental model has been developed and fabricated by the Moscow Design Bureau for Cine Equipment. In the folded form it is the size of a book. Reading is accomplished through

an eyepiece, which is a drawback of the equipment. The enlargement is from 6 to 40 times (Fig. 24). This device is equipped for reading microphotocopies on roll film of 16 and 35-mm width or microcards. Insertion of the microcards for reading is convenient and is accomplished very quickly. Reading may be accomplished using daylight or with illumination from a desk lamp. Built-in illumination is provided using a miniature bulb supplied from a pocket flashlight battery. This device may be used to read microphotocopies in both transmitted and reflected light.

GRAPHIC NOT REPRODUCIBLE

Fig. 20. Microfot reader.

GRAPHIC NOT REPRODUCIBLE

Fig. 21. Attachment to Microfot reader for reading aperture punchcards.

<u>The Mark VII reader of the Microcard Corporation (USA)</u> is intended for reading microphotocopies on an opaque base in reflected light (Fig. 25). The magnification factor is 24. The screen size is 250 × 300 mm. The device has a convenient arrangement for inserting and shifting the microcard during reading, is equipped for operation in a lighted room; a fan is provided to remove the head from the illuminating lamp.

GRAPHIC NOT REPRODUCIBLE

Fig. 22. AChS-1 reader.

GRAPHIC NOT REPRODUCIBLE

Fig. 23. AChM-2 reader.

GRAPHIC NOT REPRODUCIBLE

Fig. 24. AChK-3 reader.

GRAPHIC NOT REPRODUCIBLE

Fig. 25. Mark VII reader for reading microcards on an opaque base (in reflected light).

GRAPHIC NOT REPRODUCIBLE

Fig. 26. Filmac-100 reader-copier.

GRAPHIC NOT REPRODUCIBLE

Fig. 27. PES reader-copier.

For efficient production of optical (enlarged) copies from the microphotocopies being read use is made of reader-copiers which are a combination of a reader with a device for rapid copying. The rapid production of optical copies is based on the corresponding photographic processes using specialized photographic materials.

The Filmac-100 reader-copier of the firm Minnesota Mining and Manufacturing (USA) has been produced since 1958. The enlargement ratios are 7, 9, 13, 19, and 26. The screen size is 240 × 265 mm (Fig. 26). The time for obtaining a single optical copy is 3 sec [15].

The model PES reader-copier of the Rekordak firm (USA) has been produced since 1962 (Fig. 27). The screen size is 200 × 300 mm. The enlargement ratio for reading is 23, for printing 21. The time for producing a single copy is 1 minute.

The Diophot M-35 reader-copier has been produced by the Barton Mount (USA) firm since 1962. The enlargement ratio is up to 15. The screen size is 575 × 430 mm. The time for producing a single copy is no

more than 25 sec. The electrostatic method of producing copies is used.

REFERENCES

1. A.S. Tauber, W.S. Myers. Photochromic Micro-amges-a Key to Practical Microdocument Storage and Dissemination, Paper presented at "National Microfilm Assoc.,"Conf. Washington, D.S., 1962 April, 25-27.
2. J. Kosar. Thermal development of diazotype papers. – Photographic Science and Engineering, 1961, 5, No. 4, pages 239-243.
3. B. Friedland. Technological Advances in the Thermal Diazo Process – Photographic Science, 1962, 10, pages 174-177.
4. W.A. Seifert, W.F. Elbrecht. The Kalvar Process – Photographic Science and Engineering, 1961, 5, No. 4, pages 235-238.
5. G.H. Davison, Microcards and Microfiches. – The Library Association Record, 1961, 63, No.3, pages 67-78.
6. S.A. Lipkin, G.G. Vorob'yev, L.P. Shchegolev. Modernizatsiya traditsionnykh kartotek s primeneniyem mikrofotokopirovaniya [Modernization of Traditional Card Files with the Use of Microphotocopying]. Nauchno-tekhnicheskaya informatsiya, 1963, No. 7, pages 9-11.
7. S.A. Lipkin, G.G. Vorob'yev. Izgotovleniye i chteniye aperturnykh perfokart. [Preparation and Reading Punch Cards]. Nauchno-tekhnicheskaya informatsiya, 1963, No. 3, pages 40-46.
8. Special report. Companies are devising new techniques for combating what is coming to be known apprehensively as the nation's "technical information crisis." – Chemical and Engineering News, 1961, July 17, pages 103-114.

9. Automatic selection and reproduction of industrial documents – Data Processing, 1959, Ocotober-December.

10. A. minicard system for documentary information. – American Documentation, 1959, October.

11. R.L. Laurent. Magnacard. – Data Processing, 1962, 4, No. 10, pages 33-38.

12. B.E. Bushor. Information storage retrieval. – Electronics, 1962, June 29, pages 40-61.

13. A new microcite machine. – Technical News Bull., 1961, 45, No. 7, pages 113-115.

14. P.D. Bradshaw. The Walnut system; A large capacity document storage and retrieval system. – American Documentation, 1962, Nc. 3, 270-275.

15. National microfilm association. Guide to microreproduction equipment. Edited by Hubbard W. Ballou, second edition, 1962.

Manu-script Page No.	[Transliterated Symbols]
33	м = m = mikrofotokopiya = microphotocopy
33	д = d = dokument = document
35	НИКФИ = NIKFI = Nauchno-issledovatel'skiy kinofotoinstitut = = Scientific Research Institute of Cinematography
37	ГОСТ = GOST = Gosudarstvennyy obshchesoyuznyy standart = State Standard
39	КИНАП = KINAP = kinoapparatura = cine equipment
41	МКБК = MKBK = Moskovskoye konstruktorskoye byuro kinoapparatury = Moscow Design Bureau for Cine Equipment

44 ленкинап = lenkinap = Leningradskiy zavod kinoapparatury = = Leningrad Cine Equipment Plant

50 ВИНИТИ = VINITI = Vsesoyuznyy institut nauchnoy i tekhnicheskoy informatsii = All-Union Institute of Scientific and Engineering Information

60 Мосгорсовнархоз = Mosgorsovnarkhoz = Sovet narodnogo khozyaystva Moskovskogo gorodskogo ekonomicheskogo rayona = Economic Council of the Moscow City Economic Region

72 КЕМЗ = KEMZ = kinoelektromekhanicheskiy zavod = cine and electromechanical plant

USE OF EDGE-PUNCHED CARDS FOR ABSTRACT INFORMATION ON THE BASIS OF THE RZh METALLURGIYA

T.P. Kolesnikova and A.V. Shmeleva

The publication of the abstract journal "Metallurgiya" beginning in 1953 has made it possible to concentrate on its pages the metallurgical literature which is randomly scattered in the scientific and technical journals, collections, and works in which there is no precise specialization and the articles are published in any of these publications at the whim of the author and the editors. Thus, in the RZh Metallurgiya during the five-year period beginning with 1956 publications have been covered from 3921 titles of periodical and serial editions alone. We can consider that the RZh Metallurgiya provides quite complete information on the foreign and Soviet material published in acceptable time periods. It has become considerably easier to follow the entire current world literature on any question of metallurgy. Figure 1 presents the basic indices which characterize the quantity, volume, and time that the materials are in the metallurgy section of VINITI.

This consolidation was absolutely necessary for the organization of a unified flow of information in a given branch of science or engineering. But the summary volume of the RZh – a tremendously bulky and expensive publication – could be acquired only by large organizations. In very few libraries could the reader find all the sections of the RZh Metallurgiya. The following step toward facilitating the search for current information on metallurgy was the division in 1958 of the summary volume of the RZh Metallurgiya into several issues on individual

fields, clearly demarcated with respect to subject matter. In 1958 there were only two issues on the production of iron and steel and on welding. In 1964 the number of such issues increased to ten: theory of metallurgical processes; metallurgical heat engineering; control and measurement instruments and automation of metallurgical production; production of iron and steel; metallurgy of nonferrous and rare metals; rolling and wire drawing production; welding; metal science and heat treatment; technical analysis in metallurgy; technology and equipment for casting production; technology and equipment for forging-stamping production. At the present time the last two issues do not appear in the combined volume of the RZh Metallurgiya, but in the combined volume of Tekhnologiya mashinostroyeniya [Machinery Technology]. Figures 2 and 3 show the growth by years of the number of subscribers to the combined volume and the abstract sections on metallurgy. We see from the figures that the efficiency of the utilization of the abstract information has increased sharply.

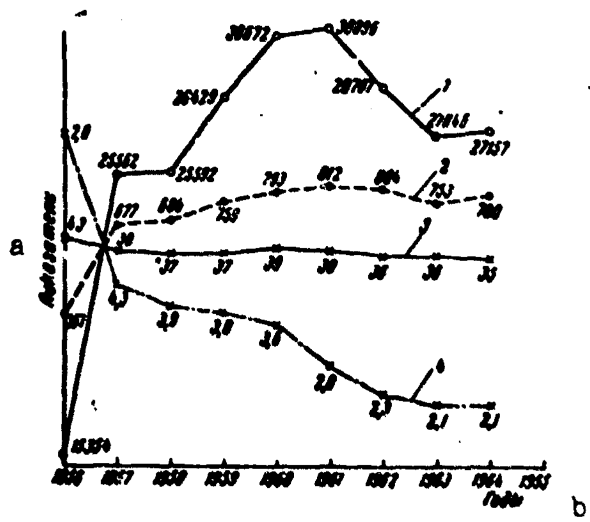


Fig. 1. Indices of the RZh Metallurgiya. 1) Number of abstracts in combined volume per year; 2) size of combined volume per year, in avtorskiy list; 3) density, in abstracts per 1 avtorskiy list; 4) average time for processing materials in the metallurgy section, in months. a) Indices; b) years.

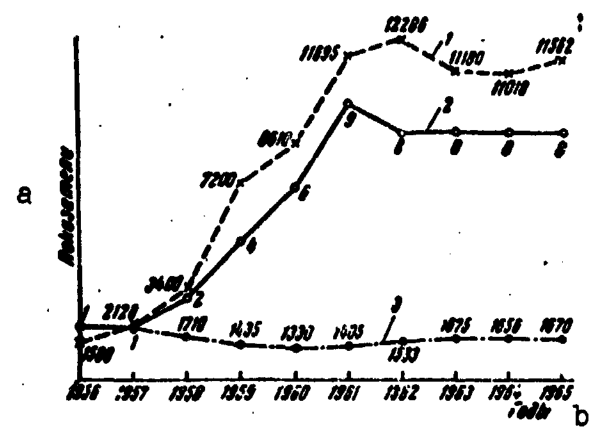


Fig. 2. Circulation of the RZh Metallurgiya and its issues by year. 1) Total circulation of combined volume and its sections; 2) number of sections in the combined volume of RZh Metallurgiya; 3) circulation of combined volume. a) Indices; b) years.

But the problem of the tremendous excess of information in the combined volume, somewhat reduced in the sections, for each specialist has continued to remain unresolved.

If we follow the path of further fractionation of the sections of the RZh, then this will inevitably lead to an increase of information loss, since with the present technique for preparing the RZh at VINITI the duplication of abstracts in the various sections of a single combined volume is eliminated.

The intimate interrelationship of the branches of science and engineering makes it very difficult to centralize all information in issuing the sections of the RZh, since in practice it is impossible to duplicate abstracts in all those sections in which this would be required for a given specialist.

But the primary deficiency of RZh is in retrospective search. This search with the use of the index is quite complicated, slow, and does not permit the use of any mechanization. Scanning of the subject indices for several years, usually issued with some delay, writing out the numbers of the abstracts with subject matter close to that desired,

finding the necessary copies of the RZh, and the abstracts in each RZh, and the subsequent digging out of the original sources — this is the far from simple way in which the information search is conducted at present.

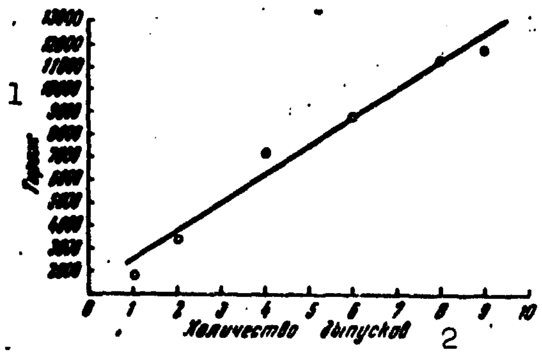


Fig. 3. Variation of circulation with number of sections published in parallel with the combined volume of the RZh Metallurgiya. 1) Circulation; 2) number of sections.

The referral nature of the subject index along with the abstract nature of the headings leads to considerable information noise during the search. The present breaking down of the headings of the subject indices makes them ever more cumbersome. A survey of the scientific workers in the field of metallurgy has established that the search for required material using the annual subject indices to the RZh, as a rule, does not reduce the time in comparison with search through the RZh without an index.

All these circumstances indicate the serious necessity for the development of different information search systems which will make it possible to accelerate and facilitate the retrospective search, provide a multi-aspect approach in each given search and in time, and, finally, the creation of individual and specialized information reference sources.

We considered that it was necessary to make more efficient use of the basic RZh Metallurgiya with an improved process for systematizing

and processing the primary information. Edge-punched cards answered our requirements most completely.

In this connection at the VINITI in the metallurgy section with the aid of the Scientific Council on Cybernetics of the Academy of Sciences USSR there was developed in 1963 and published in 1964 9 (10 in 1965) punchcard catalogs on metallurgy using the standard K5 punch-card format with two-row edge perforations.*

The type of card which we selected permits arranging on it a complete abstract from the RZh Metallurgiya, which to a considerable degree eliminates the referral nature of the card file. The number of perforations intended for coding the subject and bibliographic labels is 200, quite adequate for any specialist.

In developing the card prototype we attempted to create a standard prototype which would be clear, simple in operation, and basically could be used for abstract information in any branch of science and engineering. The standardized model makes it considerably easier to add to the information reference store abstracts from any related fields.

The left side of the internal field of the card and the corresponding right side on the reverse were set aside for the abstract. The right side remains free. On special order a window may be cut here and a microphotocopy of the given publication mounted (up to 40 pages). Thus, by locating the text on the free field of the punchcard, we obtain a very convenient microlibrary with the possibility of information retrieval. There is no necessity for storing either specialized or abstract journals. By elementary means the reader extracts only those cards which are obviously of interest to him.

It was decided to assign the great majority of the locations (about 150 and any combination of them) for coding the subject labels which reveal the contents of each publication.

For clarity on each card there is printed a stencil with notation of the subject and bibliographic labels using coded symbols grouped in individual fields. The stencil shown in Fig. 4 is common to all the card files of the RZh Metallurgiya and in the future will be extended to related fields.* The code symbols in the left, right, and lower margins on each card file have their own particular meanings.

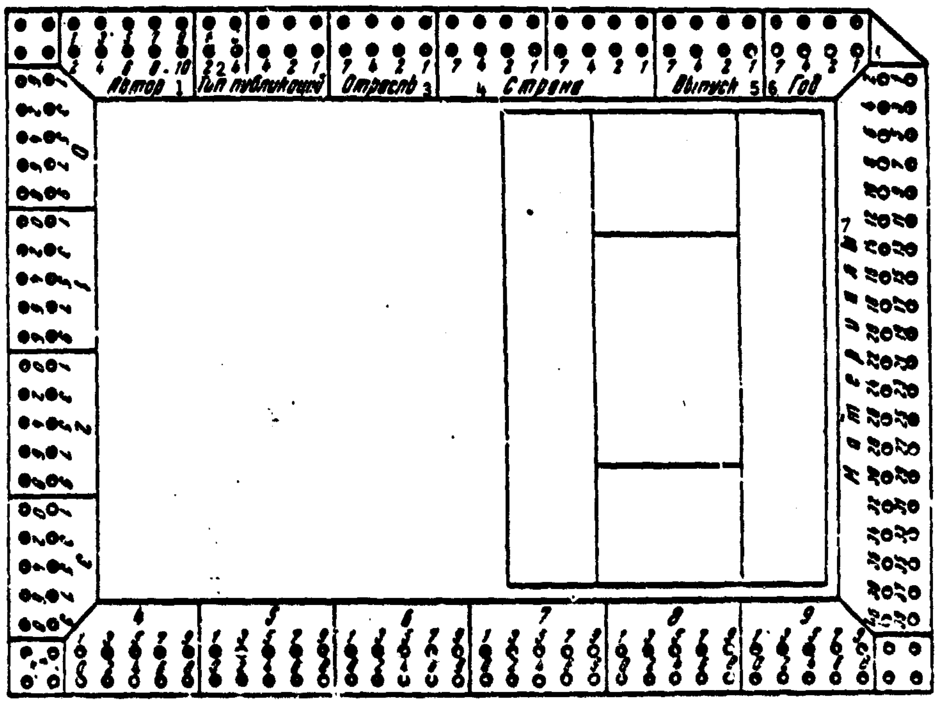


Fig. 4. Punchcard model. 1) Author; 2) publication type; 3) field; 4) country; 5) issue; 6) year; 7) materials.

The upper margin of the card is intended for the following labels (from left to right): surname of author (first letter), publication type, branch of industry and country on which information is presented in the publication, number of the card file, year of publication.

The subject labels of each card file, grouped using the local decimal system, are coded on the left and lower margins of the card. Each of the ten groups is assigned a field with 10 openings, numbered using Arabic numerals from 0 to 9. The zero field and the zero punch of each

group are intended for coding general questions of a given card file or a given group of labels. The last field and the opening in each field designated by the digit 9 are reserve, intended for coding questions which with respect to subject matter relate to the given card file or given group of labels, but which were not reflected in the originally developed system. Each search subject label has its own index. For example, the index II 2.1 denotes: IV is card file IV (Nonferrous metallurgy), 2 is the label (field) 2 (Smelting), 1 is the sublabel (Tower smelting).

The local decimal system is very convenient, since it is simple, clear, easily understood by the people who code the information document during its publication and by those who use it. It is not difficult to memorize all ten basic headings of the card file, clearly differentiated with respect to subject. In this case it is easy to detect the need for shifting the information document to a different card file. This system makes it possible to express clearly in the search language the multi-aspect nature of the metallurgical literature. Excessive expansion and extension of the hierarchy of the search language would divide the attention of the one doing the coding and would create additional difficulties in searching. Both these factors, as is known, lead to loss of information.

The right margin of the card is intended for coding the chemical elements of the periodic system, metallurgical production, and other materials (40 positions).

Both combined and direct coding methods are used in the card catalogs on metallurgy. The combined code is used to encode the bibliographic mutually exclusive labels, and the direct code is used for the parallel subject labels and materials. In the first case, to each bibliographic label corresponds a combination of two or more definite

notches. In the first case, to each label corresponds a single definite type of notch: a shallow notch, corresponding to an opening of the outer row; a combining notch (slit) is a label corresponding to an opening of the inner row, a deep notch indicates both labels together. In searching for a label denoted in the outer row of openings, the needle is inserted in the outer opening and the cards with shallow and deep notches fall out, i.e., all the cards containing the desired information. When searching for a label corresponding to the inner row of openings, the needle is inserted in the inner opening and the cards with a deep cutout fall out; the stack of cards remaining on the needle are grasped with the left hand, the needle is removed and inserted in one of the corner openings – the cards with a slit then fall out.

Along with the possibility of coding the contents of each publication with respect to all the headings of a given card catalog the principle of direct coding of subject labels with the aid of the shallow and slitted notches provides completeness and the absence of information noise when extracting cards from the array.

We did not set ourselves the task of creating a search language which would completely provide for information search with respect to very narrow subject questions. The functions of our headings are more distributive. The decimal system which we suggest facilitates the creation of a search language necessary for the highly specialized organizations or individual subscribers.

The process of preparing the punchcards for publication begins in the bibliographic group of the metallurgy section, whose workers attach to each incoming publication in the section a working card made from plain paper, without openings, but precisely duplicating the punchcard format which has been adopted, and make on this card indications of the bibliographic labels in ink: author, year of publication, type of pub-

lication with respect to form. A working number is immediately written on the working card. At the stage of scientific or control editing of the abstract mark on the working card all the labels which reflect the contents of the given publication.

Then the abstracts are gathered under the headings of the RZh Metallurgiya into an issue, are sent to the production printing combine of VINITI and are printed. A single typesetting is used for printing the abstracts in the journal and on the punchcards. At the locations on the punchcards corresponding to the code marks in ink on the working card typographical marks are made — shadings. The notches are made by the user himself in accordance with the suggested or his own local decimal classification scheme.

The vacant combinations in the codes, the unused openings in the fields, and the free fields are used for supplementing the card catalog system in the process of usage. The three corner perforations themselves are not used, since they perform an auxiliary role in searching and permit the extraction from the card file of erroneously inverted cards. When the classification is extended, the subject headings are given new values by partial or complete changeover, and for combined storage a signaling notch is made in one of the unused corner perforations of the new cards. In the case of a slight alteration of the model all the cards with obsolete notches are removed from the card file, the notches are patched or new ones are made.

It should be noted that since the first year of publication of cards with edge perforations we have increased the number of subscription issues based on the RZh Metallurgiya from 9 to 63. The average number of subscribers to each set had doubled by 1965. Thus, the total circulation of the RZh Metallurgiya with its sections and the sets of punchcards is about 18,500 copies in 1965.

The further fractionation of the subscription sections will be determined by the profitability and the organizational capabilities for their distribution.

An analysis of the operation of the reference information sources based on the K5 abstract cards will be made after some time has passed.

Manuscript Page No.	[Footnotes]
81	This work was performed under the guidance of Candidate of Technical Sciences T.P. Kolesnikova.
82	Since 1966 this model in a somewhat modified form has been accepted as mandatory for all abstract information on metallurgy.

Manuscript Page No.	[Transliterated Symbols]
77	РЖ = RZh = referativnyy zhurnal = abstract journal
77	ВИНИТИ = VINITI = Vsesoyuznyy institut nauchnoy i tekhnicheskoy informatsii = All-Union Institute of Scientific and Engineering Information

WAYS OF USING INFORMATION SEARCH SYSTEMS IN MEDICINE

N.M. Amosov and O.P. Mintser

Medical science is undergoing great difficulties in connection with a whole series of aspects. The modern ideas on the physiological and pathological processes within the organism are unusually varied and include the vast knowledge accumulated over the centuries. Thousands of journals, collections, books are published daily devoted to the results of the work of thousands and thousands of doctors and researchers.

The selection of journal articles for recent years on a particular subject can be made only by continuous scanning of all the journals of this specialty. However, we can never be certain that we have not missed something important, and the existing bibliographies, in addition to the fact that they are far from complete, are published with a considerable delay. Hence the research results become accessible for a wide circle of doctors only when their timeliness has to a considerable degree been lost in connection with the appearance of new methods of study and treatment. This is why the treatment institutions prefer to duplicate studies rather than find their results in the literature.

On the other hand, considerable difficulties in finding and processing information in clinics are associated with the numerous methods for treating patients, which in the majority of cases are not reflected in the appearance of new forms of documentation. The records in histories of disease frequently differ in their extremely disconnected nature, the absence of systematization, and frequent errors. All this makes the search for the required information difficult.

In medicine, as in any other science, a definite situation is characterized by a considerable group of data. This leads to the necessity for continuous search for new correlation connections. A simple table for 100 given studies, let us say, of 100 patients, with the aid of which we can determine the nature of a few combinations of characteristics already leads to definite difficulties. With increase of the number of conditions in the group the possibility of establishing relationships is practically lost.

Only a few years ago the narrow specialization of scientists aided in the solution of problems of data reduction. Actually, in some small field it is easier to find the information of interest, and observations in specialized hospitals of a homogeneous group of patients made it possible to draw certain conclusions. But today narrow specialization offers only partial aid, since attendant illnesses and complications may be missed.

An essential limitation in information search and processing is the absence of a unified classification of diseases, complications, and patho-physiological shifts with respect to the various organs and systems. This hinders the exchange of information between the treatment institutions and prevents an improvement of the precision of clinical medicine.

Of considerable assistance in overcoming these difficulties is a new branch of knowledge which has been termed documental information system theory. The study of this subject has become a science in itself, and primary attention is devoted to questions of creating information search systems – IPS. The application of these systems in clinics, considering the possibility of subsequent entry of the data into a unified medical information center, may yield results in the very near future whose importance it is difficult to overestimate.

First, the IPS aid in overcoming the gap between the outstanding and small clinics, between the specialists and the general practitioners, which is the result of the possibility of using the experience of all the treatment institutions.

Second, we can overcome a considerable portion of the subjectivity associated with the inadequate qualification of doctors in evaluating the information obtained.

Third, the IPS aid in the process of establishing a diagnosis, when the question arises of what minimal amount of research (which the clinical is capable of conducting, of course) must be carried out with respect to the patient so that the diagnosis will be most reliable, while the studies themselves will present the least risk, and the technique for their conduct will be sufficiently simple. This means the practical implementation of the problem which has long faced medicine — a maximum of data with a minimum of analysis. The urgency of this last question does not require any comment — it suffices to note that in the clinics tens of analyses are frequently conducted, and a correct and reliable diagnosis is still not obtained.

There are many versions of the information search systems, which in principle are divided into the following 4 types: 1) punchcard; 2) using microfilms; 3) based on the use of electronic computers; 4) combined systems.

Most widely used for processing comparatively small arrays of information are the punchcard systems, while the digital computers are used for processing large amounts of information. We shall not discuss the history of creation of IPS, we simply note that in foreign medical practice systems for accumulating, storing, and retrieval of information are widely used. In the Soviet Union the information search system was first developed at the A.A. Vishnevskiy Institute of Surgery in

Moscow. Work is proceeding on the design of clinical IPS in Leningrad, Minsk, and other cities.

However, the programs for finding the most efficient system are being carried out too independently. Therefore, in the Thoracic Surgery Clinic of the Ukraine Scientific Research Institute for Tuberculosis and Chest Surgery work is being done on developing a unified plan for studies in this direction. As the first stage it is proposed that the elements be developed for a new approach to the evaluation of clinical information, and we should shift over to this approach immediately.

1. STANDARDIZATION OF METHODS OF STUDYING THE PATIENT AND EVALUATING THEIR RESULTS

The existing techniques for conducting clinical examination of patients are already out of date. Each treatment institution performs these studies using its own plan, even if the series of tests are the same in scope, the evaluation of their results frequently differs. Without even discussing the various forms of recording, this hinders the comparison of data obtained by the clinics. Closely related to this situation is the question of objectivization of the methods of examination and the standardization of equipment.

Thus, a unified scheme for conducting analyses, a generally accepted gradation of the results, and their evaluation have become an urgent necessity.

2. CREATION OF UNIFIED CLINICAL CLASSIFICATIONS

The necessity for this work is a result of the fact that certain clinics continue to divide a particular illness into three stages, others use four stages, a third group may not classify the disease at all.

In addition, at the present there is no strict systematic classification of complications, there is no clarity in the definition of patho-physiological shifts with respect to the organs and systems for different diseases. Finally, there are no criteria for evaluating the efficiency of the diagnostic and treatment methods.

3. STANDARDIZATION OF MEDICAL DOCUMENTATION

Within the clinics the creation of unified schemes for documentation is going forward in two stages, first, transfer to punchcards (or to any other material information carrier) of the data of periodic laboratory and biochemical analyses, doctors' notations, etc. This is the so-called primary documentation (first level of the information system).

Second, the transfer of the entire volume of information contained in the history of the illness in a compressed historical information form (second information level). An example of the organization of such an information system will be given below.

Even these steps alone considerably improve the degree of precision of clinical medicine, particularly in specialized hospitals with a large number of similar patients. However, in the unspecialized or small treatment institutions this is obviously not enough. It is necessary to create a medical information center.

Academician V.M. Glushkov has played an important role in the development of the idea of creating an information center.

The problems here are the following:

1) the development of a standardized classification of diseases together with the outstanding specialists and scientific societies;

2) the creation of a standardized form of medical documentation for all clinics, convenient for transferring the information to the IPS:

3) collection and storage of medical information;

4) processing data from the literature and comparing them with results obtained in the clinics;

5) formulation of complex diagnoses and selection of optimal treatment methods;

6) conduct of statistical and correlation calculations;

7) coordination of the scientific studies associated with statistical processing of clinical material;

8) aid the treatment institutions in introducing the new methods of diagnosis.

We picture the medical information center as an organization having the complex structure whose schematic may be as shown in Fig. 1.

All the information collected using the unified scheme is fed either to the collection section, where it is subjected to analysis, or to the reference section, if it contains several inquiries. From the collection section the data enter the archives, and from the reference section the data enter the computer room. In the memory of the digital computer there are stored programs for retrieval of diagnostic, treatment, and prognostic information, on the basis of which a plan is formulated for seeking similar cases.

On the basis of the materials from the archives of disease history and the literature data, the digital computers provide a summary of similar cases with some degree of approximation, a probabilistic diagnosis, and an optimal treatment method. However, in the process of formulating the diagnosis a version is possible in which considerable additional information is required for a more precise diagnosis. Then the request for this additional information is sent back to the treatment institution. The procedure is similar in the treatment process, when the continuously changing symptoms lead to the need for constant and

continuous refinement of the treatment method and require a determination of the possible prognosis of the patient's condition.

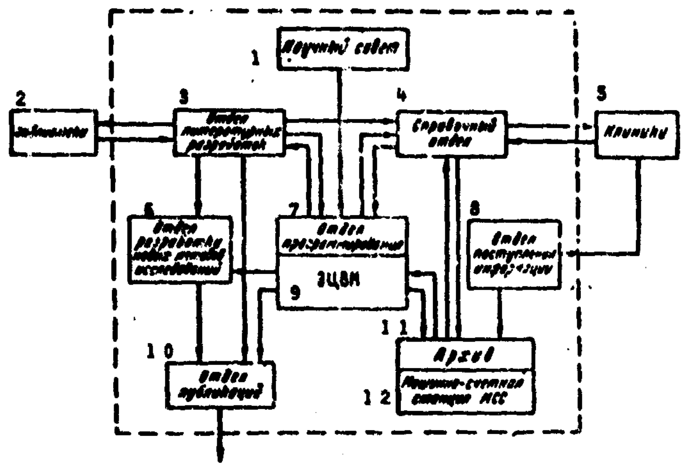


Fig. 1. Schematic of medical information center. 1) Scientific council; 2) libraries; 3) literature processing section; 4) reference section; 5) clinic; 6) section for developing new research methods; 7) programming section; 8) information input section; 9) ETsVM; 10) publication section; 11) archives; 12) computer station MSS.

It is clear that the information material will be transmitted over communication channels both in the form of digital data and in the form of curves. A division of the Institute of Cybernetics of the Academy of Sciences of the Ukrainian SSR is working on the possibility of such transmission.

In the archives, where information from the clinic and literature data enter, use must be made of the latest systems for accumulating, storing, and retrieving the required information. The vast system memory is the result of the large amount of data, while the necessity for retrieval of information in many cases in the course of a short time interval requires quite high selection speed of the IPS. Finally, the frequent referral to the system and the need for outputting all possible copies of the originals requires the use of copying and duplicating equipment in the information center.

The second part of the archives, intended for recording and accu-

mulating literature materials, uses the data from the abstracting section. The latter will be in communication with all the libraries and abstracting centers of the country and will exchange information with foreign institutes. In addition, use will be made of the work of the scientific researchers of the various medical institutions.

The medical information center is a facility in which a large amount of factual and literature material is accumulated. Information concerning the latest researches and treatment methods come here first. Finally, the center has varied equipment for processing data. Therefore it is advisable that the medical information center publish communications based on the considerable clinical and literature material.

From these same aspects it becomes obvious that the information center must offer consulting and procedural assistance to the clinics in developing and verifying new methods. In addition, it may also perform coordination functions. The latter is particularly important in connection with the fact that in spite of all the methods of reporting the number of duplicative studies (which involve, moreover, a tremendous number of researchers) increases every year.

The formulation of programs for medical information search, the evaluation of new methods of diagnostics and treatment will be particularly valuable if they are performed by specialists. Therefore the participation of outstanding scientists in the information center is desirable.

The principle of gradualness must be observed in organizing the information service. First small information search centers are created within the clinics. These information systems will first be concerned with individual nosological units or groups of diseases. At this stage of the work we can make an evaluation of the research methods, determine approximately their reliability, and consequently identify the

most important and definitive methods.

A unified information center is created on the basis of a group of such systems. As the second step there must be statistical processing of the characteristics and a determination of the different versions of the course of the illnesses. This makes it possible to find a clinical precedent with adequate degree of accuracy.

The accumulation of clinical data permits determining for the surgical clinics instructions for operations, the risk of operative intervention, and the condition of the patient in the near and distant post-operative period. Later on it will be possible to formulate a disease program, which constitutes the third stage.

It is obvious that in connection with changes of medical viewpoints, objectivization of the research data, the appearance of new diagnostic methods and treatment methods a periodic review of the archives will be required, first within the clinical systems, and then in the medical information center. Free space for entering modified and additional information must be provided for this purpose on the information carrying material.

As an example let us examine the construction of an information search system implemented in the previously mentioned Thoracic Surgery Clinic. Its general arrangement is planned in the following form (Fig. 2).

The primary information is all the data obtained in examining the patient. This information is fed periodically into the information search system. For example, for each day of the post-operative period we propose to use a single machine-sorting punchcard (the first level of the information system discussed above). With the aid of this card catalog we will be able in the future to determine the versions of the course of several illnesses. After release of the patient the entire

history of the sickness will be entered in the IPS, from which data are selected for the second level of the system. This selection is performed with the aid of specially developed epicrises. They record the entire complex of investigations introduced into the system for information search, information groups, and indices. The groups are gradations of numerical quantities. In the process of their compilation at first literature information is used, after which they are modified using the data from our clinic by means of preliminary statistical processing of 100 histories of illnesses and, finally, are discussed with the leading specialists. The information indices are the record of the columns and positions of the machine punchcards corresponding to the given characteristic.

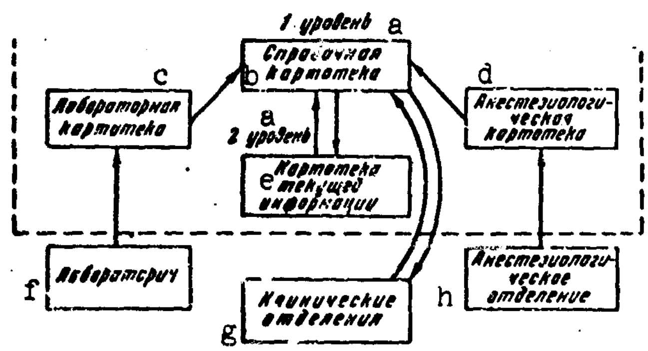


Fig. 2. a) Level; b) reference card file; c) laboratory card file; d) anaesthesiological card file; e) current information card file; f) laboratories; g) clinical sections; h) anesthesiological section.

Thus, in the epicrises there are located complete data on the patient and the corresponding information groups are recorded. All this permits obtaining the required information from the epicrises and permits the verification of the correctness of entry of the data onto the punchcards. The information indices are written out by technical personnel.

A complex of computing and perforating machinery is used in the clinic. Its application is justified, first, by the frequency of refer-

ral to the archives, the rapid and continuous growth of the archives; second, by the necessity for conducting computer operations; third, by the desire to then introduce data from the machine punchcards into the electronic equipment. At the present time archives of a history of illnesses involving heart defects is being developed. The volume of information from each history exceeds 1500 characteristics. They are arranged on three punchcard models. The first is intended for recording the data of the pre-operational period, the second is devoted to the operational period, while the third records the characteristics of the course of the illness in the post-operative period.

The punchcards are correlated by the number of the illness history and by the year that the patient entered the clinic, and are differentiated, first, by the color, which makes it possible to separate them visually, and second, by the specially coded first digit of the illness history number, which makes it possible to perform machine selection of the cards.

On each punchcard space is left for entering additional information which is the result of studies which are not mandatory. The reverse side serves for a brief record of unforeseen characteristics of the course of the illness. Finally, on each punchcard there are recorded the conclusions from the two other cards relating to the given patient, in particular the diagnosis, the success of the operation, and its immediate results. The latter makes it possible to perform an independent processing of the punchcards.

Preference is given to direct coding, which facilitates decoding of the material and the retrieval. However, in several cases, for example in the part of the second punchcard concerning the hemodynamics of the operational period, and in the third punchcard, devoted to the changes of the symptoms in the post-operative period, use is made of

the combined type of coding. In addition, on the second and third punchcards use is made of a special scheme for the characteristic of the symptom variations: the magnitude of a characteristic was within definite limits in the post-operative (operational) period. If an increase (decrease) is noted, a remark is made of the level of the change, in what stage it occurred, whether it occurred once or several times, and for how long.

The anaesthesiological records will also be entered in the IPS of our clinic. They are recorded on manual punchcards, a prototype of which is being developed. On their unused field there will be located graphs of the hemodynamic variations, and along the borders there will be information on the use of medications, the nature of the operation, characteristics of the narcosis, etc. In addition, there will be a grid drawn in which the number of the column of the machine punchcard corresponding to the recorded characteristic is noted. The latter is done for convenience in transferring data to the machine punchcards.

A similar card file is also being created for the clinical laboratories, however in this case machine punchcards will be used.

Finally, we note that each punchcard is duplicated so that one copy is located in the sickness history records, a second in the archives, and the third in a spare card file.

4. CONCLUSIONS

The following operations may be performed with the use of the system described above.

1. Search for a clinical precedent. The solution of this problem permits at the very first stages of the system operation the determination of the risk of operative intervention, the modification of a diagnosis, and evaluating the possible severity of the patient condition in

the post-operative period.

2. Verify the reliability of several methods of research.

3. Perform statistical and correlation calculations.

4. Determining the efficiency of treatment methods.

Naturally, the suggested scheme is still far from perfect, the widest possible range of doctors and researchers of the scientific and technical information institutes must participate in its development. However, the results of the use of the information search systems in medical practice will not be long in being felt and even at the very first stages they will aid in improving the precision of clinical medicine.

Manuscript Page No.	[Transliterated Symbols]
88	ИПС = IPS = informatsionno-poiskovaya sistema = information search system
92	ЭЦВМ = ETsVM = elektronnaya tsifrovaya vychislitel'naya mashina = electronic digital computer
93	МСС = MSS = mashinna-schetnaya stantsiya = computer station

SOLUTION OF SOME INFORMATION LOGIC PROBLEMS OF GENERAL CHEMISTRY ON ELECTRONIC DIGITAL COMPUTERS

A.L. Seyfer

In the present paper we shall consider four basic trends in the creation of information logic systems for general chemistry. These are the automatic information search for chemical compounds on the basis of their physical and chemical properties; the accumulation, processing, and retrieval of information using composition-property diagrams; nomenclature automatic translation; and the information logic methods of calculating properties of compounds.

We shall describe primarily the investigations made at the Laboratory of Electrosimulation of the Institute of Scientific Information of the Academy of Sciences USSR. An extensive review of the present state of the question is presented in [1].

I. AUTOMATIC INFORMATION SEARCH FOR CHEMICAL COMPOUNDS USING THEIR PHYSICAL AND CHEMICAL PROPERTIES [2]

In this section we shall consider very briefly the experience in use of electronic digital (computing and informational) machines for creating a machine information search system using the physical and chemical properties of individual (primarily nonorganic) compounds. Sometimes this system is termed an "electronic handbook." The creation of such systems is becoming more urgent with every year which is associated, naturally, with the very rapid growth of chemical information.

Generally speaking, each concrete or even potentially possible

(even including all chemistry) array of chemical information may be considered from the formal point of view as some universal class V. The elements of this class will be all properties of all chemical compounds and any combinations of them (solutions, melts, reactions, etc.), and also all conceivable characteristic conditions of determining the properties of each chemical object. The formulas of the chemical compounds, and also their systematic and trivial names and composition-property diagrams for the systems of chemical compounds are also elements of the universal class, just as all the rest. However, the obvious specific nature of such elements forces us to consider them separately. This applies most of all to the structural formulas of the chemical (primarily organic) compounds, for which the methods of machine treatment are described in detail in [3]. The questions of nomenclature translation, and also the machine recording and processing of the composition-property diagrams will be discussed in the corresponding sections of the present paper. Here we shall deal with the subclass V', including in itself all those elements of the universal class which relate only to the properties of the individual chemical compounds and the characteristic conditions for determining these properties.

The properties of the inorganic compounds are collected in the electronic handbook described below. Here we shall arbitrarily consider inorganic compounds to be those which have in their composition no more than two carbon atoms in a single radical or ligand. Of all the classes of chemical compounds it is primarily these inorganic compounds (and of the organic compounds only the hydrocarbons) which have been studied relatively completely. Therefore in the general mass of chemical compounds they stand out by the relative completeness of the collection of properties which have been studied. This, incidentally, to a considerable degree stimulated the selection of these compounds as the objects

for the information search system.

In describing substances, modern chemistry uses approximately 300 physico-chemical properties. We have selected 120 of the most important properties: thermochemical (heat and temperature of phase transitions, formation energy and entropy, etc.), thermophysical (vapor pressure, solubility, density, viscosity-temperature relations, etc.), structural (bond lengths and angles between them), optical (refraction, polarization, etc.), magnetic, electric (dipole moments, conductivities, etc.), etc. It should be noted that the set of properties may be expanded not only by using the physical and chemical constants, but also using information on the adaptability, biological activity, crystalline structure of a particular compound, etc.

From the formal point of view all the constants may be divided into three types:

a) binary criteria ("yes" – "no"), such as "does a given substance dissolve in benzene," etc.;

b) numerical characteristics, i.e., the intrinsic values of particular constants;

c) the coefficients of the equations describing the variation of the magnitude of some property with variation of the conditions, for example the coefficients of the equations of the type

$$C_p = a_0 + a_1T + a_2T^2$$

or

$$\lg P = A + \frac{B}{C+t}\,.$$

where C_p is the heat capacity at a given temperature; P is the vapor pressure of the compound; t and T are the temperatures in degrees C and K; A, B, C, a_0, a_1, a_2 are empirical coefficients.

As a rule, for the constants of the type "a" and "b" there are not

always standard characteristic conditions of definition and therefore in many cases these constants must be specifically specified, i.e., together with the magnitude of the property we indicate the conditions under which it was determined. Each formal type of constant requires its own methods of recording and its methods of processing during search. We should not one specific characteristic of the formulation and organization of an information array on the physical and chemical properties of compounds which is associated with the fact that for many (particularly the more simple and popular) compounds in the literature we encounter several values of a particular constant which differ from one another. Therefore, not taking the responsibility for the selection of a particular so-called recommended value of the constant, we must present all the values, specifying each time the literature source.

1. Forms of Search Strategy

In the automatic information search system for the inorganic compounds there are defined three types of questions which may be specified or, as is more frequently said, three types of search strategy.

1. Find in the machine memory and printout one or several definite properties of a specific compound.

2. Find in the machine memory all the properties of a given compound and print them out. In this case, if for a given compound there are also recorded the coefficients of an equation, it is necessary to calculate the values of the property defined by this equation with a given step of variation of the argument and print out the entire table or a portion of the table. In the case when certain constants have somewhat differing values, it is necessary to print out all these values with an indication of the literature sources.

3. Find in the machine memory and print out all the compounds hav-

ing a certain specified set of properties. For example, all the compounds containing sulfur but not containing chlorine, having a dipole moment which is nonzero, melting in the limits 150-180°C, boiling no higher than 270°C, etc. Such a type of search strategy is most suitable for the practical problems of selecting a material with given properties.

2. Organization of the Information Array in the Machine Memory

In spite of the fact that the inorganic compounds have been studied more than the others, only 7-8% of all the possible constants or the entire array are available. The remaining 92-93% constitute the so-called void constants. The latter are explained primarily by the still inadequate state of investigation of particular compounds and, to a lesser degree, by the inapplicability of a given type of constant to a given compound (for example, the index of refraction of an aqueous solution of a compound which is not soluble in water, etc.).

Therefore, the simplest, tabular system of recording the constants, where each property of each compound is assigned its own memory cell, is not really applicable. In this case the required memory volume must be increased by a factor of 12-13, which naturally leads to approximately the same increase of the search time. This last situation is associated with the fact, as will be shown below, that the processing time for any information subarray in the operational machine memory is incomparably less than the time for entry into this operational memory of the succeeding subarray from the external memory units: magnetic tape, punched cards, etc. However, the increase of the information volume leads to a nearly directly proportional increase of the total time for input of the information from the external memory units.

Naturally, the primary problems of organization of the information

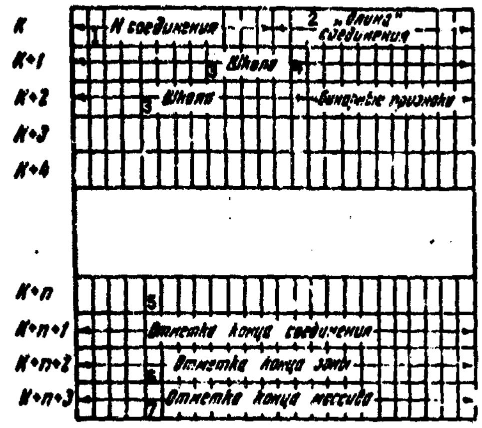


Fig. 1. 1) *N* compounds; 2) "length" of the compound; 3) scale; 4) binary characteristic; 5) end-of-compound mark; 6) end-of-zone mark; 7) end-of-array mark.

array in the machine memory are the problems of information compression, the struggle against the previously mentioned void constants. When operating with the universal computer which was first utilized the information array was organized so that a separate section of the machine memory was assigned to a single compound. In places 1-12 of cell κ there is recorded the number of the compound, and in places 13-24 of this same cell there is entered a number equal to the number of cells assigned to the given compound. The cells $\kappa + 1$ and $\kappa + 2$ are assigned to the scales. A "1" in a definite in a definite position of the scale cell indicates that a property value is recorded for the given compound. The number of the place in the first scale cell or the number of the place in the second scale cell plus 24 (24 is the machine word length) corresponds to the number of the property in the previously fixed list of 38 properties. The last 10 places of the $\kappa + 2$ cell are assigned to the binary characteristics (is the compound soluble in C_6H_6, C_2H_5OH, CCl_4, and so on).

The values of the physico-chemical properties are stored in cells

$\kappa + 3, \kappa + 4, \ldots, \kappa + n$. In this case, if the value of some property is missing in the array, the cell does not remain empty, but in it there is stored the value of the following property, etc. Thus, if in the scale cells there are ones only, for example, in 2, 6, 16, 24, 31 (7 + 24) positions, then the values of the properties corresponding to

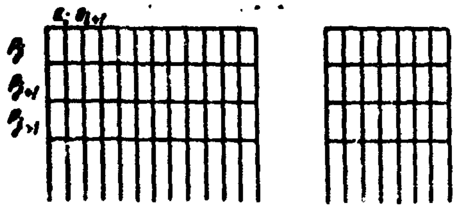


Fig. 2.

these positions are recorded in sequence in the same order and occupy only 5 cells in place of 38 for the case of the tabular form. The cells $\kappa + n + 1$, $\kappa + n + 2$, $\kappa + n + 3$ are assigned for the marks "end of compound," "end of zone" (there are 3-4 compounds in a zone), and "end of entire array."

In the Ural-4 machine the information array concerning the physico-chemical properties of the inorganic compounds is organized somewhat differently. The long machine word length permitted replacing the scales by a peek-a-boo matrix. The latter is a line-by-line simulation of superposition punched cards [4]. The structure of the peek-a-boo matrix is shown in Fig. 2. Here a single definite compound is placed in correspondence to each identical position of a group of cells (column). Each cell (matrix row) corresponds to a single definite property. Thus, in a single subarray, "packet," there are gathered 40 compounds. The number of properties in the set is not fixed ahead of time. A portion of the matrix cells is assigned for the binary characteristics. Ones in a definite position of any cell indicate the presence in the memory of a value of the property p_i of the compound a_j. The values themselves

of the properties are recorded in the same fashion as in the case when the scales are used. First the constants of the first compound are recorded sequentially, then the second, etc. After the values of the properties in the individual cells there are written the addresses of the subarrays with the equation coefficients. At the end of the array there are written, symmetric to the values of the constants, references to the source from which this constant was taken, and after all this there are written the addresses of the so-called duplicates of the constants, i.e., the second, third, ... values of certain constants (with their literary references), for which the first values are presented in the basic part of the subarray.

TABLE 1

TeO

001+2+471000	1069
006+2+365000	10297

Since on the Ural-4 machine only a single packet is stored on a single zone of the magnetic tape, there is no need to indicate the number of the cells assigned to a compound or to specially number the compounds. The number of the compound is automatically determined by the number of the packet, and within the packet it is determined by its place in the packet. On the average a single packet (40 compounds) occupies approximately 500-700 cells.

We use the method of preparing the array with storage. It is intended for preparation of information for a set of alphanumeric perforators with the objective of subsequent input into the Ural-4 machine. As the information arrives it is entered on cards for perforating. A sample of one card is shown in Table 1. On it there are written out and coded the formation energies for gaseous (property 001) and

liquid (property 006) states of the compound TeO together with the literature citations (1069 and 10297). These cards are accumulated constantly, independent of the perforation process. During perforation only a single property of a single compound is entered on each punchcard.

Before entry into the Ural-4 machine, a selection is made on a conventional sorter of all the punchcards relating to a single compound. The array which has been subsorted in this fashion can, in addition to its basic use for input to the Ural-4, pass through the printer unit of a tabulator and obtain a summary of the physicochemical properties of all the compounds accumulated over some definite time period.

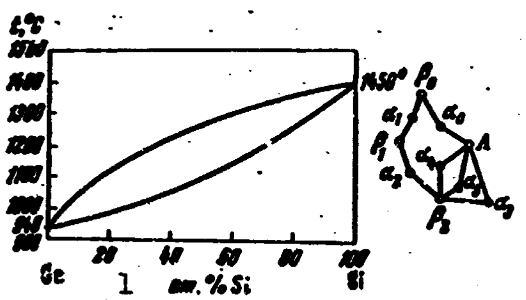


Fig. 3. The Ge-Si system and its physico-chemical graph. 1) at.

The translation of the array from the input language to the machine language is accomplished automatically by specialized translation programs. The basic idea in these programs is the conversion of the punctuation form in the input languages to a position code and primarily the formation of scales or a peek-a-boo matrix. The description of the translation program is the subject of a separate study.

3. Search

The search programs are formulated so that, depending on the interrogation, i.e., on the type of search strategy, particular blocks of

the unified search program are activated. The blocks of this program are the following: processing of the peek-a-boo matrix, scale shift, comparison of the value of a property with the limiting values specified by the question, and calculation using the equations.

The block for matrix processing operates with a series of questions, each of which is essentially the disjunctive normal form of the binary characteristics (property numbers). The result of the block operation is a table of scales — TO, each scale of which contains the numbers of the compounds in a packet which have the given property (regardless of its numerical value).

The second block is intended for finding in the information array the required numerical values of the constants from the address generated by the preceding block. The third block compares these numerical values with the limiting values specified by the question and, finally, the last, fourth block calculates the values of the constants specified by the equation coefficients. Activation of a particular block is accomplished by a control subprogram depending on the type of search instruction.

Experimental operation of the search program on a machine with the scale form of information array for 3-4 compounds in one zone gave the following results. Using the first and second types of search instruction about 2 min is necessary for an answer to a single question (not counting the time for input of the question itself). This time is made up of the zone search time — about 80-90 sec, the search in the zone — 3 sec, and the time for printing the result — 30-40 sec. Using the third type of search instruction for each zone the times were: 20-25 sec for entry of the next zone, 6-8 sec for searching in the zone, and 5-10 sec for printing the result of the search for a given zone. Thus, in all the processing of a single zone occupies 31-43 sec, which for

all the material of inorganic chemistry would give about 30 hours. Naturally, this length of time for a continuous scan of the entire array is excessively long and makes the entire system of little use in practice.

However, the use of the peek-a-boo matrix on the Ural-4 machine permitted a significant reduction of the search time. Thus, for the third type of search instruction ("find all the compounds having a definite set of properties") complete processing of a single packet together with the time necessary for entry into the operational memory of the succeeding packet of 40 compounds does not exceed 5-6 sec. A continuous sort of the entire array and the search in the array for the required compounds for all inorganic chemistry is accomplished in 20-25 min. However the search using the first two types of search instructions lasts 20-50 sec on the Ural-4 machine.

Here we should note that using the described scheme we can construct automatic search systems for other than the chemical compounds. We can assume that the metals and their alloys, biological, economic-statistical, and other objects may serve as input information for such a system. An attempt has recently been made to use an electronic handbook for creating an information search system for the metals and alloys [5].

II. INFORMATION STORAGE, PROCESSING, AND RETRIEVAL USING THE COMPOSITION-PROPERTY DIAGRAMS

A somewhat unique field of application of the information-logic methods of chemistry is the information concerning systems of substances, i.e., the field which is studied by physico-chemical analysis. Modern chemistry deals ever more frequently not only with the individual chemical compounds, but also with their combinations - systems. Alloys

of the metals and salts, solutions, etc., are very widely used by engineering, which has led to the need for their systematic study by the methods of physico-chemical analysis. This in turn has given rise to and is continuing to give rise to the accumulation of data concerning the equilibrium chemical systems in ever increasing scale.

Just as in certain other branches of chemistry, physico-chemical analysis has its international language — the geometrized language of the composition-property diagrams. The composition-property diagram itself, according to the definition of N.S. Kurnakov, "is a graphical representation of that complex function which defines the ratio between the composition and the properties of homogeneous bodies or phases which are formed in a system" [6]. Figures 3-7 show several specimens of such diagrams. The accepted technique of representing in the form of certain geometric patterns the variation of some physical property as a function of the physico-chemical interactions in a system as a function of the variation of the composition has led to a tendency to treat certain typical topological structures on the diagrams as one-to-one representations of the physico-chemical transformations which take place in the systems. Seemingly, such a tendency is quite fruitful if only because it is essentially an attempt at some standardization and formalization of all the accumulated material. However, such standardization and formalization, unfortunately, is not all-encompassing, since in actuality there is not a complete one-to-one correspondence between the physico-chemical transformations in the system and their "typical" geometrical representations. This implies that the geometrical structure of the diagram in itself cannot be considered as complete information concerning the system. Therefore, we [illegible]t consider all three aspects of the system (and the diagram which represents it) as inseparably related: the topological, metrical, and chemism (chemical

essence) aspects. Consequently, the machine language of physico-chemical analysis must take into account these three aspects of the form of the information concerning the equilibrium systems and at the same time with respect to its capabilities must be equivalent to the real language of physico-chemical analysis. By machine language of the physico-chemical analysis, we mean here a definite system of recording in the memory and logical processing by the machine of the experimental and theoretical information with the objective of solving information problems. In the real language of physico-chemical analysis the following typical information problems are solved.

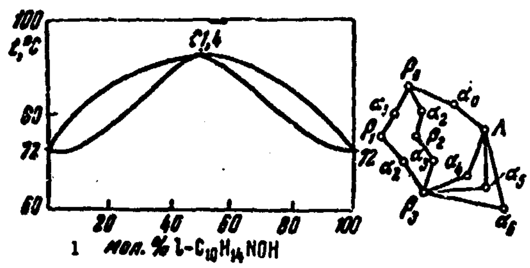


Fig. 4. The levo-carroxime/ dextro-carroxime system and its physico-chemical graph. 1) mol.

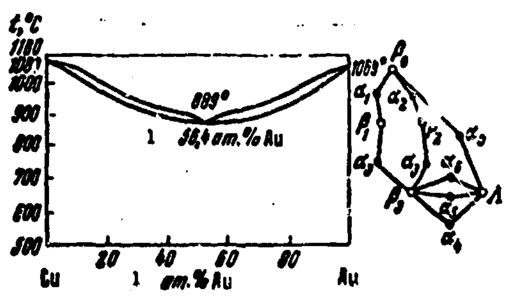


Fig. 5. The Cu-Au system and its physico-chemical graph. 1) at.

1. Finding a concrete composition-property diagram (for example, the diagram of the states of the Si-Ge system or the electrical conductivity-composition diagram of the system H_2SO_4-H_2O).

2. Finding all the diagrams for a given system of substances (for

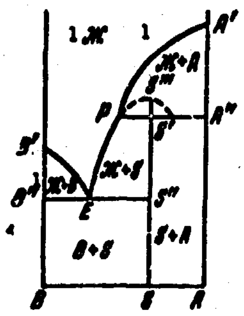


Fig. 6. Diagram of the condensed state of a binary system with complete solubility in the liquid state and complete insolubility in the solid state, with the formation of an incongruently melting chemical compound. 1) Zh.

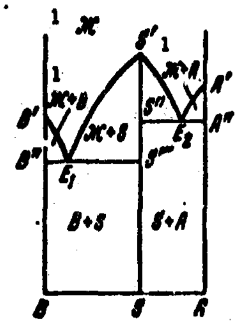


Fig. 7. Diagram of condensed state of a binary system with complete solubility in the liquid state and complete insolubility in the solid state, with the formation of a chemical compound which is not dissociated in the liquid state. 1) Zh.

example, all the composition-property diagrams: density, electrical conductivity, saturated vapor pressure, refraction, etc., for the system $K_4[Fe(CN)_6]-H_2O$).

3. Finding the diagrams for systems in which one or more of the components are in the general class of the chemical compounds (for example, all the state diagrams for alloys of the Mg chlorides with the chlorides of the alkali metals or the composition-electrical conductivity diagrams for the systems H_2O-normal $C_nH_{2n+1}OH$).

4. Identify systems with a common structure for the entire diagram or for individual segments of the diagram (for example, liquid systems with a maximum of the vapor pressure or condensed systems with extrema of the initial melting temperatures).

5. Determine the property magnitude for a given composition for one or more systems of substances (for example, the crystallization

temperature of all alloys containing 27 at. % Zn or the melting points of the alloy 82% Fe and 18 wt. % Ni, etc.).

6. Find the entire combination of substances with definite properties or with a definite combination of properties (for example, the composition by components of mixtures of all salts melting above 320°C without decomposition and having a density higher than 1.7 g/cm^3, etc.).

These same information problems must also be solved in the developed machine language of physico-chemical analysis. Let us consider the three forms of description in machine language which correspond to the three aspects of the real language of physico-chemical analysis. The description principles described below have been developed basically for the simplest and at the same time most frequently encountered case of binary systems. The majority of these principles are also applicable to the description of three-component and multi-component systems.

Topological description. N.S. Kurnakov introduced into physico-chemical analysis the topological representations, i.e., the consideration of the composition-property diagrams as a closed topological complex. In describing topological structures by binary composition-property diagrams it is necessary to take into account the basic characteristics of the contemporary electronic machines for information processing – they all require that the information be represented in linearized form for input, i.e., that the information consists of a linear sequence of symbols in machine code. It is obvious that the composition-property diagrams for the binary and particularly for the three-component and multi-component systems are essentially nonlinear. Thus, any diagram of a binary system regardless of its chemical significance is a two-dimensional closed complex, composed of points, lines, and surfaces (fields). Because of the nonhomogeneity of such an initial complex its representation in linearized form is associated with over-

coming certain difficulties. When separating from this complex a reference graph consisting only of lines (edges of the graph) and points (vertices), information will be lost concerning the fields and their relative positions in the diagram since for such a graph several different plane realizations are acceptable.

For the two-dimensional topological complexes, these difficulties (in particular, for the binary system diagrams) may be avoided by using the duality principle and introducing "coloration" of the vertices, and also special "boundary-vertices." The graph constructed with the use of these techniques and uniquely describing the composition-property diagram for the binary systems has been termed a physico-chemical diagram [7]. The sequential restructuring of the original graph into the physico-chemical graph may be followed in Fig. 8, where as the original complex we have taken a typical diagram of a system with a eutectic point. First we construct a graph which is dual to the original. To do this a point of the field-vertex of the dual graph *A* (Fig. 8, A) is correlated with each field of the diagram. The entire region lying outside the closed complex forms an empty field, to which the empty vertex *A* of the graph is correlated. The graph *A* thus obtained describes only the relative positioning of the fields in the original diagram. After constructing the dual graph each intersection of the edges of this graph with an edge of the original graph is replaced by a point, a "boundary-vertex." In the newly formed graph *B* (Fig. 8, B) the boundary-vertices are located between the field-vertices. For each field of the diagram their number is equal to the number of lines separating the given field from the others. Therefore, for example, the field lying below the solidus line in Fig. 8 is described by the single field vertex *ß* and by five boundary-vertices *α*, which causes the appearance of new edges in the graph. By definition the vertices of type *α* will always be vertices

of only second order, two-ray. The type β vertices will have an order no less than 2, part of them will be nodes, i.e., vertices with order ≥ 3. Thus, in the graph *B* there are vertices of two types, two colors. In graph theory such constructions are normally termed graphs with labeled vertices. In certain cases, for the description it is convenient to color each vertex in its own color, which permits in a single representation having information concerning the general structure of the diagram and concerning its concrete chemical content. In this case the vertices of type α, which are formed from a single line of the original diagram, always have the same color, as shown in Fig. 8, C.

A B C

Fig. 8. Sequence of construction of a physico-chemical graph. A) Graph which is the dual of the original; B) graph with vertices of two colors; C) physico-chemical graph.

There are several methods of linear representation of graphs. We have used one of the simplest – the construction of a star table.

Let $\alpha_i, \alpha_k, \alpha_s, \ldots, \alpha_p$ be the boundary-vertices, and $\beta_q, \beta_s, \ldots, \beta_z$ the field-vertices of a physico-chemical graph. Then each row of the star table appears as follows:

$$\beta_q - \alpha_i \beta_k,\ \alpha_{i+1} \beta_l, \ldots,$$

i.e., following each symbol of any field-vertex there are written successively all the branches of the graph emanating from this vertex. Each branch consists of two vertices: the nearest vertex is type β, and that lying between two such vertices is type α. Essentially each row of the table is the symbolic representation of one definite field of the

diagram with an indication of the neighboring field and the bounding lines. The complete set of rows for all vertices of the physico-chemical graph type may be written linearly. For example,

diagram without extrema of the melting points:

$$\beta_0 - \alpha_0\Lambda, \alpha_1\beta_1 | \beta_1 - \alpha_1\beta_0, \alpha_2\beta_2 | \beta_2 - \alpha_2\beta_1, \alpha_4\Lambda, \alpha_5\Lambda, \alpha_3\Lambda |,$$
$$\Lambda - \alpha_0\beta_0, \alpha_4\beta_2, \alpha_5\beta_3, \alpha_3\beta_2 |;$$

diagram with a maximum of the melting points:

$$\beta_0 - \alpha_0\Lambda, \alpha_2\beta_2, \alpha_1\beta_1 | \beta_1 - \alpha_1\beta_0, \alpha_2\beta_3 | \beta_2 - \alpha_1\beta_0, \alpha_3\beta_3 | \beta_3 - \alpha_6\beta_1,$$
$$\alpha_3\beta_2, \alpha_4\Lambda, \alpha_5\Lambda, \alpha_6\Lambda | \Lambda - \alpha_0\beta_0, \alpha_6\beta_3, \alpha_5\beta_3, \alpha_4\beta_3 |;$$

diagram with a minimum of the melting points:

$$\beta_0 - \alpha_0\Lambda, \alpha_2\beta_2, \alpha_1\beta_1 | \beta_1 - \alpha_1\beta_0, \alpha_2\beta_3 | \beta_2 - \alpha_2\beta_0, \alpha_3\beta_3 |,$$
$$\beta_3 - \alpha_3\beta_3, \alpha_2\beta_2, \alpha_6\Lambda, \alpha_5\Lambda, \alpha_4\Lambda | \Lambda - \alpha_0\beta_0, \alpha_6\beta_3, \alpha_5\beta_3, \alpha_4\beta_3.$$

In practice for the initial input of the physico-chemical graph for further processing on a computer the symbols of the vertices of type α and β usually are numbers obtained by arbitrary numbering of the lines of the original diagram.

With further processing (canonization) of the diagram these numbers are replaced using a special program by symbols denoting both the phase composition of the fields and the chemism of the phase transitions (lines) in the system. Here we must point out that when coding the search strategy all questions associated with the search for general or concrete topological structures must also be represented in the form of a description of some physico-chemical graph. For example, such a graph for the solid solutions will be $\alpha_i\beta_j\alpha_k$, where β_j is a nonnodal vertex. Then the finding of the answer to such a question will be the search using a special program for the embeddings of our subgraph (question) in all the graphs described in the machine memory.

<u>The analytic description</u> of the diagram is a representation of its metrics. In a narrow, concrete sense the metrics of a chemical diagram is the relationship between the magnitude of a property and the compo-

sition of the system. This relationship is reflected either graphically on a diagram constructed from a table of experimental data or analytically with the aid of some equation. Numerous attempts have been made to find theoretical (based only on the properties of the original components and obtained by interaction of the components of the compounds) relationships for the lines of the composition-property diagram which lead to equations describing qualitatively the interactions in the system, but which do not offer the possibility of performing quantitative calculations. On the other hand, even Mendeleyev in his studies on solutions solved successfully the problem of approximating experimentally the resulting curve on the composition-property diagrams by equations of parabolas. Among the recent studies in this field, of interest are the investigations of Reshetnikov, who has very successfully approximated the isotherms of the binary systems by expressions which in the final analysis are reducible to an equation of the form

$$y = \frac{a_1x^2 + a_2x + a_3}{a'_1x^2 + a'_2x + a'_3},$$

where y is the magnitude of some system property; x is the content of one of the components, %; a_i, a'_i are coefficients which are determined empirically by the method of least squares [8].

Therefore we have used as the basis for the analytic description of the chemical diagrams of binary systems the representation of the chemical diagram as a combination of lines expressed in analytical form [9-11]. In this case each line of the diagram reflects the functional relationship $y = f(x)$, where y is the value of a property of the binary system of substances of a given composition, x is the percentage of one of the components. As we have already mentioned above, so far there are no theoretically derived relationships for all the cases of the property lines. Therefore there remains only one approach — the conversion

of tabular data or data taken from a curve into some empirical formula. This is essentially the replacement over a limited interval of the complex functional dependence of the system properties on the properties of the original components by a relatively simple empirical formula.

First we tried the Lagrange interpolation polynomials. In spite of the relative simplicity of the interpolation process itself, the memory volume occupied by the data for interpolation is quite large and the analytical form of the diagram constructed on the base of the interpolation formulas is extremely complex. However the primary shortcoming of the interpolation formulas in application to the chemical diagrams is their low accuracy.

Thus, the Lagrange interpolation polynomial

$$y = \sum_{k=1}^{n} y_k \prod_{i=1}^{m} \frac{x - x_i}{x_k - x_i}$$

for one of the property lines of the state diagram of the Bi-Te system at individual points gives a deviation up to 20-25% between the calculated and experimental values.

Incomparably better results are given by approximating the property line by some polynomial using the method of least squares. In selecting the approximation method and the form of the approximating polynomial we must take into account two to some degree contradictory factors: the simplicity of the process and the accuracy of the results obtained. On the basis of universality considerations we chose a conventional power polynomial of the form:

$$y = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0.$$

In the approximation using this polynomial on a digital computer we took into consideration the following factors.

1. The degree of the approximating polynomial was not established ahead of time, but was worked out automatically each time on the basis

of a given maximal value of the deviation. For the state diagram, following Gromakov [12], we took

$$\varepsilon = \frac{T_{max}}{256},$$

where T_{max} is the maximal value of the temperature on the given curve.

2. All points, including those relating to the pure components, are of equal value. The arguments that on the state diagrams the points corresponding to the temperatures of the phase transitions of the pure components are absolutely reliable quantities while all the remaining points are considerably less reliable seems to us to be in error. Numerous facts of revision of individual constants of the individual substances are the best proof of this. In any case, there are no bases to consider that in a given series of experiments the parameters of the substances involved are identical to the parameters of a specially purified and precision measured compound.

3. The approximation is performed using the method of least squares with solution of the required systems of linear equations by the Gauss method.

The first experiments showed that with this method of approximation it is easy to achieve the required accuracy and the resulting polynomials rarely have a degree greater than 4-5. For the usual smooth curves without essential extrema (nearly all the lines on the diagrams are of this type) normally the 3*rd* degree is adequate. Thus, for the analytic description of the diagram there are entered into the machine memory for each line the coefficients of the polynomial in decreasing powers of the variable and the abscissas of the end points of the curve, i.e.,

$$l_i - a^i_n, a^i_{n-1}, a^i_{n-2}, \ldots, x^i_1, x^i_2.$$

The system described above for representing the metrics of the di-

agram requires having during implementation search algorithms in the machine memory and some program for calculating the polynomial. We took for this purpose the simplest Horner scheme which may be used to calculate the values from a given value of x. However, the Horner scheme is not suitable for such search problems as the determination of the property-composition from a given quantity. Therefore, it was necessary either to supplement the analytical description by programs for the approximate solution of high-order equations or to expand the analytic representation itself, introducing into it polynomials which are inverted relative to the variables. We investigated both of these approaches. Supplementing the analytic expression itself required a double approximation, i.e., for each property line we introduced into the memory two polynomials:

$$y = f(x);\ x = \varphi(y).$$

This doubles the time for the initial machine processing of the diagram and doubles the length of the analytic expression. However, the search using this expression is accomplished with the aid of the same Horner scheme and requires little machine time. The use of the method of approximate solution of high-order equations does not require increase of the time for the initial processing of the diagram and does not increase the memory volume used for the analytic representation of the diagram. However, during the search it is necessary to perform for each diagram relatively lengthy (12-17 sec on the Ural-4 computer) computational operations, which for an array of 1000 diagrams requires an additional 20-25 min of time. It appears that the first approach, i.e., the use of the inverted polynomials, is more efficient in spite of its complexity. This method has been used in the search program described below.

The general form of the diagram gives information concerning the

chemical composition of the components and phases on the diagram. Using as an example the binary metallic systems, an attempt has been made [13] to give for each line and each field of the diagram a standardized formal notation. Thus, for example, all the verticals of the congruently melting compounds formed by the interaction of the original components are denoted by $A_mB_n^{II}$, the lines of the eutectic transitions are denoted by AB_e, the eutectoid transitions by AB_{ed}, the synthetic transitions by AB_s, the liquidus lines by $L_1\alpha$, the solidus lines by $\alpha\alpha$, etc. In the analysis of a large number of published diagrams there were found finite sets of lines which describe (bound) the same regions. The processing of these data made it possible to compile tables of the so-called regional logic polynomials. Thus, the region L (liquid) in the general form will be expressed by the following regional logic polynomial:

$$[A_mB_n^{I} \vee \overline{A_mB_n^{I}}] \wedge [A_mB_n^{II} \vee \overline{A_mB_n^{II}}] \wedge K_3^{m_1} \wedge L_1\alpha^{m_1}$$

Thus, it is possible, at least for the binary metallic systems, to determine from the form of the logical polynomial the nature of the region and, consequently, there is no need for a special introduction into the machine memory of information concerning the regions on the state diagrams. The nature of the general representation may be seen in greater detail from the description of the experimental automatic search system using the state diagrams. The system was designed for the Ural-4 computer.

The first version of the automatic search system using the state diagrams does not take into account all the forms of the representations developed for the diagram chemism. In particular, it does not reflect the standardised formal notations of the lines which was mentioned above. Therefore the information concerning the regions and the singular points of the diagram are entered ahead of time, and are not

worked out in the process of the initial processing of the diagram.

For the initial processing all the information concerning the diagram is divided into three subarrays D_i, OST_i., Sh_i, where D_i is the subarray containing information on the diagram metrics and the chemism of the diagram regions; OST_i is the subarray containing a list of the singular points of a given specific diagram using the previously fixed list of standard singular points; Sh_i is the "cap," i.e., the general form of the diagram.

For the entire collection of diagrams, consequently, there are three individual arrays.

1. The array of diagrams

$$M_1 = N_1Д_1, N_2Д_2 \ldots,$$

where N_i is the recording depth of a single diagram (number of cells).

The initial form of D_i is constructed in turn as follows:

diagram number,

number of the 1*st* curve (using an arbitrary enumeration of the curves within each diagram),

the coordinates of the points of this curve

$$x_1y_1, x_2y_2 \quad \text{etc.}$$

the punctuation mark A_1 (end of curve 1),

number of the 2*nd* curve and so on for all the curves,

punctuation sign A_2 (end of coordinate recording of all the curves),

region number a_1,

number of the uppermost curve in the region,

number of the lowest curve in the region,

punctuation sign A_3 (end of region), and so on, for all the regions,

punctuation sign A_4 (end of region array), region dictionary; the

region number is the designation of the region,

punctuation symbol A_5 (end of region dictionary),

punctuation symbol A_6 (end of diagram).

II. The array of singular points is

$$M_2 = N_1\text{OCT}_1, N_2\text{OCT}_2, \ldots$$

The initial form of the OST_i is constructed in turn as follows:

diagram number,

singular point code,

its coordinates $(x_i y_i)$, and so on, for all the singular points of of the diagram.

III. The cap array is

$$M_3 = \text{Ш}_1, \text{Ш}_2, \text{Ш}_3, \ldots$$

Each cap includes

diagram number,

system components,

code of the property described by the diagram,

code for the dimensions of the composition expression,

bibliographic citation code.

The order of arrangement of the diagrams in these three separate arrays may be arbitrary and they may not correspond to one another. The identification is accomplished using the diagram number.

The initial processing of the diagram array is accomplished simultaneously for the entire collection and has as its objective the reduction of the initial representations to a form which is convenient for rapid and complete search. In particular, from the array M_1 there is developed a step (here and in the following in this section a step is the number of cells assigned) for each diagram, a step for the curve subarray, for the region subarray, etc., and there are also worked out by approximating the tabular data the polynomials $y = f(x)$ and $x = \Phi(y)$

for each diagram curve using an *a priori* specified magnitude of the acceptable deviation. In this case the polynomials are described by the coefficients of the variable with an indication of the degree of the polynomial.

The entire array is reorganized so that it will be easy to identify all the diagrams on the three basic arrays. After this initial processing we obtain three new arrays m_1, m_2, m_3.

The m_1 array is obtained by processing the N_1 array. The essence of the processing was described above. The m_2 array is the same singular point array N_2, while the cap array m_3 becomes dominant and is arranged as follows:

$$m_3 = Ш_1 \langle Д_1 \rangle \langle OCT_1 \rangle, Ш_2 \langle Д_2 \rangle \langle OCT_2 \rangle \ldots,$$

where $\langle D_i \rangle$ is the address of the diagram D_i in the array m_1; $\langle OCT_i \rangle$ is the address of the subarray of singular points of the diagram D_i in the array m_2.

The search program solves seven basic search problems.

Problem n_1 – for a given concrete diagram determine from the composition the magnitude of a property (from x find all values of y).

Problem n_2 – for a given concrete state diagram or for all state diagrams with some one common component generate the value of a property on the liquidus line from the given composition (from x find y_{max}).

Problem n_3 – for a given diagram using a specified magnitude of a property determine the system composition (one value of the composition).

Problem n_4 – for a given diagram or for all diagrams in an array, using the value of a property determine all the system compositions corresponding to this pr[illegible]

Problem n_5 – from the given coordinates of a describing point find the diagram region in which it is located.

Problem n_6 – determine all the singular points of the diagram.

Problem n_7 — output all information concerning the diagram.

It has been shown that the majority of problems reduce either to these seven types of basic problems or to their combinations. The question (inquiry) prior to input into the computer must have the form

$$Z = N_{\text{inquiry}} \cdot n_{\text{problem}} \cdot d \text{ (data } x_i y_i \text{, and so on).}$$

The search program permits simultaneous introduction of 5-7 inquiries (i.e., no more than 20 values of x_i and 10 values of y_i). The search itself consists of two major portions: the generation of the control sequence and the extraction of the information. The control sequence determines in what order we must refer to the arrays m_1 and m_2, and only after this is the information concerning the diagrams extracted and the search problems solved. During the search the array m_3 is transferred by parts in order to generate the control sequence from magnetic tapes onto the computer magnetic drum, while the arrays m_1 and m_2 remain on the magnetic tapes throughout the search. The problem solution time does not exceed 30-40 sec for an array with 400 diagrams.

On the average the time for the initial processing of each diagram is about 3 min.

We see from the system description that its use still does not permit the solution of problems associated with finding the diagrams of common topological structure. To do this, in the new system there must be introduced a representation of the "chemism" of the lines, i.e., a representation of the nature of the physico-chemical transformations in the system as the describing point moves over the diagram. In addition, the entire diagram must be described as a physico-chemical graph using the method of the star table. Such a system will permit solving any information search problems.

The collection of composition-property diagrams stored in the computer memory permits the solution also of certain correlation and com-

putation problems of physico-chemical analysis. In this case the computing capabilities of the machine itself are used in full measure. An example of the correlation problems is the statistical verification [14] of the Kordes equation [15]:

$$\frac{x_a}{x_b} = \frac{T_b - T_e}{T_b} \cdot \frac{T_e}{T_a - T_e},$$

where x_a, x_b, T_a, T_b and T_e are respectively the molecular (atomic) concentrations and crystallization temperatures of the components and the eutectic. Using as an example 20 metallic and 34 salt systems with a eutectic (without secondary complications) on a digital computer using the data stored in the computer memory, it was shown that the Kordes equation is not precise, and may be used only as the basis for a very rough statistical prognosis. In this case, after finding the sections of the surface described by the Kordes equation, we found the direct regression equations for the salt systems:

$$\left[\frac{x_a}{x_b}\right] = 0{,}111 + 1{,}02\left[\frac{x_a}{x_b}\right]$$

and for the metallic systems:

$$\left[\frac{x_a}{x_b}\right] = 0{,}024 + 0{,}608\left[\frac{x_a}{x_b}\right]$$

It is interesting that the correlation relations derived after 20 min are close in their form to the correlation equations for the liquid azeotropic systems [16], obtained by the authors after several months of comparisons and calculations. We see from this example the vast capabilities for studying and logical processing of the material of physico-chemical analysis offered by the system of machine recording and machine processing of the composition-property diagrams.

III. AUTOMATIC NOMENCLATURE TRANSLATION

One of the first situations where the chemical information-logic

methods are applicable in processing various chemical information is the automatic conversion of various sorts of chemical compound names into formulas in order to make it possible to store them uniquely in the computer memory. When outputting from the computer the results of the operation of the search algorithms frequently the resulting compound must be presented in the form of a name. In this case we need a unique inverse automatic translation, i.e., the translation of an equation into a name which is given in some systematic nomenclature system. In addition to these utilitarian problems, a similar machine translation of names may be used advantageously for compiling formula indices for abstract journals, handbooks, etc.

The possibility of using the modern electronic computers for converting the names of chemical compounds into their formulas is based on representing the name, on the one hand, and the formula of the chemical compound, on the other hand, as mutually unique calculable functions. With this approach the machine conversion of a name into a formula (or a formula into a name) is considered as a computational process performed using a definite program. However, the compound name is not mathematical information, therefore its machine conversion is a problem of a logical nature having much in common with the problems of automatic translation from one language into another.

The variety of the nomenclature systems in chemistry and primarily the ambiguity of the names in each system make the problem of automatic nomenclature translation extremely complex. This applies to the names of both the organic and, particularly, the inorganic compounds. The first attempts at automatic nomenclature translation in organic chemistry, in spite of the existence in this field of more or less universal nomenclature systems, because of the complexity of the translation rules, have been limited so far to individual classes of compounds or

types of names. In inorganic chemistry the experimental translations of names into formulas and vice versa have been conducted only by the author of the present paper with his co-workers [17-20]. In the following we shall examine some results of these experiments and the chemical bases for them.

First let us agree to call the translation of names into formulas the direct automatic nomenclature translation, and the translation of formulas into names will be termed inverse translation. In the field of direct nomenclature translation the names of the complex compounds were the first to be studied. The selection of the complex compounds for creating the first program for machine translation of names into formulas is based on the fact that the complex compounds may be named using several rational nomenclatures while the systematic names themselves are quite complex both with respect to principle of construction and with respect to structure. The trivial names of the complex compounds such as luteo salt, Peron salt, *1st* Reyze base, which are widely used and previously were in competition with the systematic names, do not require logical processing for machine decoding and are easily translated by simply finding in a dictionary the corresponding formula equivalent. However, such a dictionary becomes extremely bulky for the ever increasing number of trivial names in the literature and possibly, has already reached the saturation limit. Therefore as the initial information we selected the rational names using any of the accepted and suggested nomenclatures: Werner's [21] with supplements of Chernyayev [23] and Greenberg [22], the international IUPAC [24], the nomenclature suggested by the *8th* Mendeleyev Congress of the Commission of the OKhN of the Academy of Sciences USSR [25].

First, the rules for construction and the structure of the names of the complex compounds in all the listed nomenclatures were subjected

to formal logic analysis. In this process we identified typical roots and the so-called search operators which relate the roots in a single name. The identified roots and the search operators were combined into a "root dictionary."

The program consists of four basic blocks: identification of the roots, analysis, combinatorics, formula synthesis, and formula printout. In the root identification block the term is broken down into chemically meaningful roots. This division is accomplished with the aid of the root dictionary stored in the computer memory. Table 2 shows as an example certain sections of this dictionary.

The entire translation process from name input to formula printout is accomplished automatically and consists of 3500-6000 elementary logic and arithmetic operations. The difference in the number of elementary operations required for the translation of various names depends on the name length and the frequency of encounter in the name of letters which occupy the most "populated" portions of the dictionary. The average machine time for translating one name is 3-5 sec.

Following are several specimens of such a machine translation.

Input:

potassium dicyanoargentate, or

dicyano-(1^+) potassium argentate.

Output:

00212 00001 37777 00201 00001 00205 00002 37777 00001,

i.e., the formula $K_1[Ag_1(CN)_2]_1$.

Input:

hexachloroplat acid, or

hexachloro-(4^+) platinate acid.

Output:

00215 00002 37777 00202 00001 00206 00006 37777 00001,

TABLE 2

Слог 1	Ранг 2	Символ 3	Комбинатор 4	Код символа при записи формулы в память машины 5
		6 Основы		
7 -аргент-	*A*	Ag		00201
8 -плат-	*A*	Pt		00202
9 -кобальт-	*A*	Co		00204
10 -циано-	*B*	(CN)	00001	00205
11 -хлоро-	*B*	Cl	00001	00206
12 -аммин-	*B*	(NH_3)	00000	00211
13 -калия	*C*	K	00001	00212
14 -кислота	*C*	H	00001	00215
15 -сульфат	*D*	(SO_4)	00002	00217
		16 Поисковые операторы		
17 -ди-	α			00002
18 -тетра-	α			00004
19 -тетр-				
20 -гекса-	α			00006
21 -пента-	α			00005
-а- ($^+$)	β		00001	
-е- (4^+)	β		00004	
22 -и- (3^+)	β		00003	
23 и т. д.				

Remarks. Rank: A) complexing agent, B) ligand, C) outside-sphere cation, D) outside-sphere anion. The combiner is a number whose last digit is equal to the valence.

1) Root; 2) rank; 3) symbol; 4) combiner; 5) symbol code when writing formula in machine memory; 6) bases; 7) argent; 8) plat; 9) cobalt; 10) cyano; 11) chloro; 12) amin; 13) potassium; 14) acid; 15) sulfate; 16) search operators; 17) di; 18) tetra; 19) tetr; 20) hexa; 21) penta; 22) i; 23) et cetera.

i.e., the formula $H_2[Pt_1Cl_6]_1$.

Input:

pentaaminchlorocobalt sulfate, or

sulfate of pentamine-chloro-cobalt (3+).

Output:

37777 00204 00001 00211 00005 00206 00001 37777 00001
00217 00001,

i.e., the formula $[Co_1(NH_3)_5Cl_1]_1(SO_4)_1$.

Here the five-digit numbers of the type 00212, 00206 (with two initial zeroes) are the symbols of the elements or radicals; their values are given in Table 2. Numbers of the type 00001, 00006 (with four initial zeroes), given in the name itself or generated by means of com-

binatorics, are the subscripts of the symbols and brackets. The number 37777 is the left-hand and right-hand square brackets. The index 1, usually not written in the formulas, is retained for convenience of various sorts of logic operations with the formulas which are required for solving various information problems.

Far more complex and considerably less defined is the translation of the names of the conventional (noncomplex) inorganic compounds, for which as yet there is no rational nomenclature. The IUPAC nomenclature and the draft of the Nomenclature Commission of the Division of Chemical Sciences of the Academy of Sciences USSR may to some degree facilitate the translation, although the introduction of a new nomenclature is a lengthy process and the old names will continue to be encountered in the literature for a long time to come. But even with the availability of a unified nomenclature we must first of all face the fact that the translation process must be as formal as possible with minimal reliance on additional (other than the name itself) information. The nomenclature systems mentioned above, unfortunately, have several alternatives and admit differences of reading. Therefore, in the initial stage the problem of direct nomenclature translation of the names of the inorganic compounds was intentionally limited to the simple and most frequently encountered compounds. However, even in this case we encountered several difficulties associated with the indeterminacy of the names of even the simple compounds. The algorithm is described in detail in [18, 20]. Here we shall only mention that for the translation we must have in the computer memory the following:

1) the root dictionary,

2) the search operator table,

3) the control operator table,

4) a summary of the rules for transforming the names of the acids,

bases, and compounds of the thio-, oxy-, and halogenides and the chalcogenides.

In the root dictionary (containing more than 3000 entries) there are collected all the roots encountered in the names with their formula equivalents and additional information on the possible valency of the atoms or radicals in different situations (as anions, cations, in homopolar compounds, oxides, etc.). The search operators are prefixes of the type mono-, homo-, tetra-, hepta-, monovalent-, etc.

In the table of control operators there are collected operators of the type: basic-, acid-, mono-substituted-, secondary-, etc.

The translation process itself consists of breaking the name into roots with the aid of the root dictionary and tables, replacing the roots using definite rules by their equivalents, combinatorics (selection of subscripts for the symbols of the atoms or radicals), and "editing" the resulting formula (dropping, where necessary, brackets, rearranging symbols, etc.).

Even with the extensive additional information obtained after a detailed logical analysis of the Russian chemical literature, the experimental testing of the algorithm did not always yield a unique result. However, the indeterminacy and ambiguity is not a defect of the algorithm, but a result of the indeterminacy of the names themselves. Just as the chemist when he encounters the term "ruthenium chloride," without a logical analysis of the entire context he cannot way whether the discussion concerns $RuCl_3$ or perhaps $RuCl_4$, or even some other compound, similarly the program stored in the machine does not lead to a single formula in this case. This applies equally to the oxides, thiohalogenides, etc.

Normally for the purposes of storing the chemical compound in the

digital computer memory the direct automatic nomenclature translation is used. The problem of inverse nomenclature translation was formulated in connection with the necessity for output from the computer of the results of automatic search among the inorganic compounds. The result of one of the forms of search [2] is a list of compounds which satisfy an entire set of specified questions. Since in the computer memory there is stored a list of formulas of the inorganic compounds, we must output from the computer either these same formulas or, if this is not possible, the names. We can always present the formula in coded form, i.e., replacing the symbols of the chemical elements by some digital code notations. However, this is not always convenient. There is no Latin alphabet on the Soviet computers with alphanumeric printers (for example, the Ural-4 computer). Therefore, it was desirable to translate the formula automatically into the systematic name, which can be printed out by letters in the Russian alphabet.

The problem of inverse automatic nomenclature translation of the inorganic compounds is less associated with such a narrow, utilitarian objective. It is of interest primarily because it yields the possibility of studying from the point of view of formal and semantic uniqueness any nomenclature systems. Of no little importance in such a study is the question of minimizing the number of logic operations required. The fewer the logic operations in the translation from the formula to the name, the more efficient, the more rational the nomenclature. In the formal analysis of the nomenclature systems we must also pay attention to their universality, i.e., to the ability to encompass any inorganic compounds.

As a result of a study using the inverse automatic nomenclature translation of the IUPAC and Commission on Nomenclature of the OKhN of the Academy of Sciences USSR nomenclatures it was found that they are

not universal, since they encompass only the primary types of chemical compounds. Moreover, at times the translation requires serious reanalysis, which is associated with a considerable complication of the programs. Thus, for the names of the heteropoly anions and the heteropoly cations we must seek the central atom, calculate how many acid groups remain in the compound, etc. However, very frequently, as, for example, in the cases $K_2Mg_2V_{10}O_{28}\cdot 16H_2O$ or even $K_6MnMo_9O_{32}$ we are not able to determine the formal characteristics of the hetero- and poly anions, and the computer is forced to give two or more types of names. Attempts to improve the IUPAC nomenclature and its reflection in the draft of the Commission on Nomenclature of the OKhN have not been successful. The artificial technique of considering any compound simply as combinations of ions (anion and cation parts), used in the IUPAC nomenclature, in addition to its unjustified artificiality. complicates the programs so much that it makes the translation itself extremely time-consuming and worst of all ambiguous. The results of the formal analysis of the nomenclatures will be presented in a separate communication in greater detail. Considering the inapplicability of the existing nomenclature systems as name systems at the computer output, we have used the Nekrasov system which is most universal and absolutely unique, also as a simple set of translation rules. The essence of this system is that in the majority of cases the nomenclature translation is transformed into the operation of "putting into words" the formulas without their logical conversion. The sense of this putting-into-words is easily seen from the examples:

Na_2O – disodium oxide;

CaO – calcium oxide;

OPCl – oxo-phosphor-chloride;

B_5H_9 – pentabor-nonahydride;

SiH_2FCl - silicon-dihydro-fluor-chlorine, etc.

For the individual atomic groupings (in our notation they appear in the formula only in round brackets) the group names are retained and used. For example,

$Fe(NO_3)_2$ - iron-dinitrate;

$S_4(NH)_4$ - tetrasulfur-tetraimine, etc.

We have not been able to find a single example of an inorganic compound formula which cannot be completely uniquely translated into a systematic name following Nekrasov. The program compiled for the Ural-4 computer contains without the auxiliary dictionaries about 1500 instructions, and the translation of a formula into a name together with printout of the results (i.e., the name itself) requires 1-2 sec.

IV. INFORMATION-LOGIC METHODS OF CALCULATING COMPOUND PROPERTIES

In the first section of this paper we introduce the concept of the universal class V, whose elements are all properties of all chemical compounds and their combinations, and also all conceivable characteristic conditions for determining the properties of each chemical object. Going still further in this direction, we must introduce still another class - the set of governing laws or relations V_1. This set consists of the mathematical equations relating both the elements within the subsets of objects Q, parameters (properties) P, characteristic conditions U, and also the elements from the various subsets of the universal set V. Many of these relations are the mathematical expression of the laws, relationships, and rules already available in chemistry, derived theoretically or determined experimentally. Certain of these equations, namely the Antoine equation and the power series for the specific heat, were utilized in the previously discussed automatic search systems for the inorganic compounds. Having available in the computer memory an ar-

ray of information on a quite large number of chemical compounds, it is quite advantageous to supplement it by certain algorithms which realize the information-logic methods of nonsearch operations with arrays.

It is of advantage to include in such information-logic methods:

1) the methods of finding a concrete value of a property if the latter is contained in the information accompanying the compound in implicit form (direct calculation):

2) methods of finding (predicting) the concrete value of a property by analogy with the parameters of other similar compounds or even of this same compound, but under other characteristic conditions (comparative calculation);

3) methods of finding (predicting) the concrete value of a property of a compound from the presence in the compound of definite structural fragments (strukturno-additivnyy calculation);

4) methods of finding new functional or correlation relations in the information array concerning the properties of the chemical compounds (general approach).

We must immediately admit that this listing is hardly all-encompassing, since the development of chemistry, on the one hand, and experimental studies in the field of machine operations with arrays of chemical information, on the other hand, without doubt will bring to life new methods or will require essential alteration of the listed methods.

Let us examine in greater detail all these methods from the aspect of the basic premises and procedures for their realization on modern electronic computers.

The direct calculation methods are applicable in the general system to the degree that there exist and are described in the literature various empirical formulas of a nonadditive nature. There are many such

formulas, but they all have extremely limited field of application. For our purposes the formulas of greatest interest are those of the type $P = f(U)$, describing the variation of the property (parameter) with variation of the characteristic definition conditions. Usually pressure or temperature appears as this characteristic condition. Storing in the computer memory only the empirical coefficients of such an equation, for a given compound we always have the possibility of calculating the value of the property with variable values of the characteristic conditions. This method, as we mentioned previously, was used in the automatic search system for the inorganic compounds.

The strukturno-additivnyy calculation is based on adding some increment (partial fraction) to the individual fragments of the structural formula or even to the fractions of some measured geometric parameter of an atom or molecule. The algebraic sum of these increments and the increments of the relative positioning of the structural fragments yields the magnitude of the desired property.

This method does not work for all properties, but does for the so-called additive properties and primarily only for comparatively narrow classes of chemical compounds. The following basic equation is characteristic for each strukturno-additivnyy method of calculation:

$$P_q = \sum_{k=1}^{k} n_k p_k,$$

where P_q is the magnitude of the desired property of the given compound; p_k is the increment (partial fraction) for the given property per one definite structural fragment of the molecule or per a definite interaction of these fragments; n_k is the number of the given structural fragments in the molecule.

We used this basic equation in [26] to design a machine scheme for additive calculation of the physico-chemical properties of the paraffin

hydrocarbons as a supplement to the information search system. In those cases when in the information array stored in the computer memory there is no value of some property of the alkane (usually higher) specified by the inquiry, this machine calculation scheme is put into operation. The problem was posed of calculating the property of a paraffin hydrocarbon using not only the formula but also the name. Thus, the problem was divided into three stages: construct automatically from the name, given in some systematic nomenclature, the structural formula, identify from this formula definite structural fragments, and, finally, from the identified fragments perform the calculation. The use of the formulas and set of fragments presented in [27, 28] made it possible to create a scheme taking into account more than 10 types of properties for any paraffin hydrocarbon. For any paraffin hydrocarbon, depending on the complexity of its structure, the automatic machine calculation of the properties together with the procedure for direct nomenclature translation and identification of the structural fragments consists of 1000-2000 computer operations and lasts only 1-2 sec. This speed makes possible, first, the large-scale calculation of the constants, and, second, somewhat exceeds the speed of the search for the constants stored in the computer memory. This makes it possible to replace the search for the constants by a calculation in certain cases which do not require definiteness.

The primary drawback of this method of calculation lies in its basic assumptions. Thus, only in the first approximation can we speak of the additivity of many properties without consideration for the integral parameter of the entire molecule. Therefore the strukturno-additivnyy methods are frequently complicated by the introduction of "relative positioning" increments, for example, increments of the interaction of the structural fragments as a function of the type of interme-

diate or surrounding atoms or bonds. Essentially this complication is caused by the desire to introduce into the calculation some integral factor. The machine calculation using such complicated strukturno-additivnyy schemes has great possibilities both with respect to the set of properties which can be considered and with respect to accuracy [29].

Comparative calculation. By this term we usually mean two approaches which have a different basis. First, this is the use of a regular variation of some property with other conditions being the same in a series of "similar," "analogous," "monotypic" compounds, systems of substances, or reactions. Second, this is the use of a regular variation of some property of one substance with a variation of the characteristic conditions.

The six methods of comparative calculation formulated by Karapet'yants [30] are based on these two approaches.

The interest in the comparative methods of calculation of the properties of compounds for use in the information-logic systems is determined by the fact that they are essentially a ready-made model of comparison by analogy and operate on the most varied classes of chemical compounds. The basic difficulty in formalizing the comparative calculation methods is the *a priori* formation of series of similar compounds, monotypic reactions, etc. We made an attempt at such a formalization in [31]. However, the logical similarity criteria worked out in that study are obviously inadequate and in each particular case they must be supplemented by contensive criteria of chemical similarity. Therefore, with a very simple algorithm of the calculation itself, made using formulas of the general form

$$G_1 = aG_2 + b,$$

where G_1, G_2 are the properties or their elementary functions for one

or more objects; *a* and *b* are empirical coefficients, the basic difficulty is the technique for selecting the objects for comparison. At the present time we make extensive use of the "trial and error" method for doing this, leading in the limiting case to the sorting of nearly the entire array.

General approach. The attempt to create a regular procedure for finding the governing laws in an array of chemical information and the subsequent calculation is the so-called general (formal) approach. In developing this method we started from the postulate that to each chemical compound there is placed in correspondence some point in an n-dimensional space, where n is the number of parameters (properties) which form the basic system, and that the $(n + 1)st$ property may be represented in some closed region of variation of the parameters by a continuous function with derivatives which are bounded in absolute magnitude by the positive number M [32].

Expanding this function (not knowing its form) in a Taylor series, by the method of least squares we can find the sought governing law. Using certain structural parameters of the molecule as the basic system of parameters, we have found a series of empirical formulas for calculating the properties of the hydrocarbons [33, 34].

REFERENCES

1. G.E. Vleduts and A.L. Seyfer, Zavodskaya laboratoriya, [Industrial Laboratory], 1962, 28, 1224.
2. A.L. Seyfer, S.S. Shurova, and Yu.A. Polyusuk, In collection: "Nauchno-tekhnicheskaya informatsiya," [Scientific-Technical Information], 1963, No. 10, 26.
3. G.E. Vleduts and N.D. Stokolova, In collection: "Nauchno-tekhnicheskaya informatsiya [Scientific-Technical Information],

1963, No. 2, 23.

4. G.E. Vleduts, D.G. Lakhuti, and V.S. Chernyavskiy, In collection: "Nauchno-tekhnicheskaya informatsiya," 1963, No. 4, 19.

5. A.P. Babayeva, A.L. Seyfer., – In collection: "Nauchno-tekhnicheskaya informatsiya," 1964, No. 4, 36.

6. N.S. Kurnakov, Vvedeniye v fiziko-khimicheskiy analiz [Introduction to Physicochemical Analysis], Leningrad, 1936, 67.

7. A.L. Seyfer and V.S. Shteyn, Zhurn. neorg. khimii, [Journal of Inorganic Chemistry], 1961, 6, 2719.

8. M.A. Reshetnikov, Dokt. diss. Rukopis' [Doctoral Dissertation Manuscript]. Fondy IONKh AN SSSR, 1954.

9. A.L. Seyfer and V.S. Shteyn, Doklady na onferentsii po obrabotke informatsii, mashinnomu perevodu i avtomaticheskomu chtennyu teksta, [Reports at Conference on Information Processing, Machine Translation, and Automatic Text Reading], No. 1, Moscow, 1961.

10. A.L. Seyfer and V.S. Shteyn, Tezisy dokladov na Vsesoyuznom soveshchanii po fiziko-khimicheskomu analizu. [Reports at All-Union Meeting on Physicochemical Analysis], Izd-vo AN SSSR, 1961.

11. A.L. Seyfer, G.I. Kleynerman, and V.S. Shteyn, Information Storage and Retrieval, 1963, 1, 13.

12. S.D. Gromakov, O nekotorykh zakonomernostyakh ravnovesnykh sistem [On Certain Regularities of Equilibrium Systems], Izd-vo Kazanskogo universiteta, 1961, page 54.

13. V.S. Shteyn, In collection: "Nauchno-tekhnicheskaya informatsiya," [Scientific-Technical Information], 1963, No. 3, 31.

14. A.L. Seyfer and Z.M. Ostroukhova, Zhurn. neorg. khimin, [Journal of Inorganic Chemistry], 1962, 7, 203.

15. E. Kordes, Z. anorg. allgem. chem., 1932, A, 167, 103.

16. A.J. Johnson, J.A. Madonis. Canad. J. Chem. Engng. 1959, 37, N 2, 71.

17. A.L. Seyfer and V.S. Shteyn, Soobshcheniya laboratorii elektromodelirovaniya, [Communications of the Electric Simulation Laboratory], No. 1. Izd. INI AN SSSR, 1960, page 172.

18. A.L. Seyfer, Soobshchentiya laboratorii elektromodelirovaniya, vyp. 1. Izd, INI AN SSSR, 1960, page 183.

19. A.L. Seyfer, V.S. Shetyn S.S. Shurova., Zhurn. neorg. khmii [Journal of Inorganic Chemistry], 1961 6, 1759.

20. A.L. Seyfer, Information Storage and Retrieval, 1, 29, 1963.

21. A. Werner, Novyye vozzreniya v oblasti neorganicheskoy khimii [New Viewpoints in the Field of Inorganic Chemistry], Leningrad, 1936.

22. A.A. Greenberg, Vvedeniye v khimiyu komleksnykh soyevineniy [Introduction to the Chemistry of Complex Compounds], Leningrad, 1961.

23. I.I. Chernyayev, Izv. Instituta po izucheniyu platiny [News of the Institute for Study of Platinum], 1926, 6, 55.

24. IUPAC Nomenclature of Inorganic Chemistry, 1157, Report of the Commission of Nomenclature of Inorganic Chemistry, L., 1959.

25. Proyekt nomenklatury neorganicheskikh soyevineniy [Draft of Nomenclature of Inorganic Compounds], Izd-vo AN SSSR, 1959.

26. A.L. Seyfer, L.N. Vleduts, Izv. AN SSSR, seriya khim. [News of the Academy of Sciences USSR, Chemistry Series], 1964, No. 11, 1985.

27. V.M. Tatevskiy. Khimicheskoye stroyeniye utlevodotorov i zakonomernosti v ikh fiziko-khimicheskikh svoystvakh [Chem-

cal Structure of the Hydrocarbons and Regularities in Their Physicochemical Properties], Izd-vo MGU, 1953.

28. N.D. Litvinov, Zhurn. fiz. khimii [Journal of Physical Chemistry], 1961, 35, 1745.

29. A.L. Seyfer, and Ye.A. Smolenskiy, Zhurn. fiz. khimii, [Journal of Physical Chemistry], 1964, 38, 1288.

30. M.Kh. Karapet'yan, Zhurn. fiz. khimii. [Journal of Physical Chemistry], 1955, 29, 1328.

31. A.L. Seyfer and A.A. Zenkin, Collection: "Informatsionnyye sistemy" [Information Systems], Moscow, 1964, 107.

33. A.L. Seyfer and Ye.A. Smolenskiy, Zhurn. fiz. khimii [Journal of Physical Chemistry], 1963, 37, 2657.

34. A.L. Seyfer, Ye.A. Smolenskiy, Zhurn. fiz. khimii [Journal of Physical Chemistry], 1964, 38, 104.

Manuscript Page No.	[Transliterated Symbols]
122	э = e = evtekticheskiy = eutectic
122	эд = ed - evtektoidnyy =eutectoid
122	с = s = sinteticheskiy = synthetic
123	Д = D = diagramma = diagram
123	ОСТ = OST = osobaya tochka = singular point
123	Ш = Sh = shapka = cap
126	З = Z = zapros = question [inquiry]
127	эксп = eksp = eksperimental'nyy = experimental
127	выч = vych = vychislyayemyy = calculated
129	ОХН = OKhN = otdeleniye khimicheskikh nauk = division of chemical sciences

AN INFORMATION SYSTEM ON THE MATHEMATICAL THEORY OF EXPERIMENTATION

I. GENERAL DESCRIPTION OF SYSTEM

V.V. Nalimov, Yu.P. Adler, and Yu.V. Granovskiy

The tasks of an information system. An information system is being designed for collective search for new ideas in that field of knowledge which we might term the mathematical theory of experimentation. It is assumed that new ideas, no matter where they are stated: in studies of a mathematical nature or in experimental investigations - will aid in the development of new trends in our activity. Here we have in mind not a simple mechanical borrowing of foreign ideas, but their reanalysis. It is assumed that the information system facilitates the process of generating new ideas, at least for those investigators who have the capability of associating a large number of varied propositions.

The mathematical theory of experimentation is based primarily on the methods of mathematical statistics. But it seems to us that these two divisions of knowledge cannot be identified, just as, let us say, we cannot identify quantum mechanics with those mathematical disciplines on which it is based. In the mathematical theory of experimentation there appear many nonstatistical methods of study, but there do not appear certain deeply theoretical divisions of mathematical statistics.

This statement may be illustrated by a few examples. One of the important tasks of the mathematical theory of experimentation is finding the optimal conditions for the operation of technological process-

es. Here, along with the probability approach there is possible a strictly deterministic approach with the subsequent use of variational analysis, dynamic programming, and the Pontryagin maximum principle. In some publications both the probability and the variational methods of analysis are used simultaneously. There may also be publications on the mathematical theory of experimentation in which along with the traditional statistical methods use is made of the ideas and methods of information theory, game theory. Finally, the mathematical theory of experimentation is essentially related with the numerical methods of analysis and with modern computer technology. Obviously, the designed information system must be sufficiently selective — it must select not all works on the divisions of knowledge listed above, but only those which have direct bearing on the subject being considered.

It is planned to construct the information system so that it will permit both the following of the development of new trends in the current publications, and also carrying out a retrospective search using the publications of recent years. The first of these tasks is the result of the existing system for organizing library activities. As a rule, the libraries of our scientific research institutes and universities do not obtain even the basic foreign statistical journals. The investigators performing creative work in the field of the mathematical theory of experimentation do not have the capability of immediate analysis of the development of foreign science. The information system being designed must fill this gap. The second task formulated above is the result of the fact that the single existing system for retrospective search — the subject index of the RZh Matematika does not have a sufficiently deep indexing system. Experience shows that this information system cannot be used for precise formulation of search problems.

The information system being designed must encompass both mathe-

matical works of a methodological nature, and publications presenting the experience in application of new methods of investigation. Here account must be taken of the special place that mathematical statistics occupies among the other sciences [1]. On the one hand, this is undoubtedly a mathematical discipline whose study requires serious mathematical preparation – in all the countries of the world it is taught in the mathematics departments. On the other hand, mathematical statistics has one characteristic which is not typical of the purely mathematical disciplines. In contrast with the purely mathematical disciplines, here we cannot limit ourselves to such criteria of value as the profoundness of ideas, refinement of construction, and logical noncontradiction. In evaluating the new statistical methods of analysis we must first of all consider their practical effectiveness. Experience shows that new and very fruitful trends in mathematical statistics arise after we are able to formulate well the problem. We must be able to propose a formulation which will correspond to the intuitive understanding of the experimenters. The mathematical reanalysis of what is done by talented experimenters will aid in formulating new problems. Hence it is clear that it is very important to follow studies published in nonmathematical journals.

Search system. Considering the arguments just stated above, we have proposed two different systems for information search.

1. The search of the methodological studies will be performed by direct scanning of 10-20 basic statistical journals. Gradually, as the information system expands, the number of journals will be increased. We must be aware that with this search system we will miss all those methodological studies which are published in nonmathematical journals. Here we might mention that the International Journal of Abstracts, Statistical Theory, and Methods* performs a systematic search of the meth-

odological studies in 417 journals such as Cancer Bulletin, Forest Science, Textile Research, Tabacco Science, etc. Experience shows that methodological articles are published quite frequently in 250 specialized journals of this type. In order to fill the gaps which are unavoidable with our search system, we must enter on punched cards the abstracts from the International Journal of Abstracts, from the American abstract journal Quality Control and Applied Statistics,* and from the RZh Matematika (VINITI AN USSR). All these materials will inevitably arrive with the considerable delay which is characteristic of the abstract journals. Referral to the abstract journal must be considered as only a temporary measure. It is possible that in the future we will be able to establish a working arrangement with the International Abstract Journal in order to exchange results of the publication search before the corresponding abstracts are written and published.

2. The search for experimental studies in which use is made of the mathematical theory of experimentation must be carried out over the entire array of journals and conference publications which relate to experimental physics, chemistry, metallurgy, geology, biology, agrobiology, psychology, sociology, etc. If we limit ourselves to the most important journals, then we would still need to deal with more than 10,000 titles of various issues. In order to make this task at least to some degree feasible, we propose to conduct the publication search only with respect to the bibliographic references. From the set of publications being considered we shall select a subset in which the name of at least one of the scientists who have made a contribution to the mathematical theory of experimentation is listed in the bibliography. The first list of such scientists has been compiled. In this "key" list are included also the authors of well-known monographs and certain popular articles. With this search system there is no need to even read the ti-

tles of the articles. The search is performed mechanically, by unskilled personnel or, if this is possible, by a computer. The technique presented here has long been used in a somewhat different form by experienced personnel of the bibliographic service. When they need to obtain some idea of the contents of a publication written in some inaccessible language (Japanese, Hungarian), they usually resort to an analysis of the bibliography presented at the end of the article.

We made a little experiment on this sort of information search in the chemical and metallurgical journals. The search is accomplished on the average at a rate of 6 journals per hour, if there is direct access to the shelves with the journals. After 5 hours and scanning 30 journals we had found 18 articles of interest to us. Further experiments showed that we miss about 25% of the publications which we really need. It was found that our key list did not include certain authors whose publications are "daughters" with respect to the primary authors. It will be necessary to make a very careful study of the additions to the key list. However, the missing of even 25% of the publications of an experimental nature is not critical. In part the gaps will be filled with some delay) as a result of the abstracts taken from the abstract journals.

The search system using the bibliographic citations is used by the American firm Institute for Scientific Information, Philadelphia, USA. There are interesting examples in the information list of this firm. Two of them relate to two articles of the well-known statistician Tukey. A summary is presented of all those works in which there is a bibliographic reference to these articles. This summary is undoubtedly of great interest. By scanning these publications we can immediately obtain an idea of where the ideas presented in the article of interest to us aroused interest. A detailed analysis is presented in [2] of the

bibliographic coupling of the articles of the journal Physical Review.

Information carriers. As the information carrier in our information system we used the standard K5 punched card of 147 × 207 mm format with two-row edge punching. The sorting of these punched cards is accomplished manually or with the aid of a simple electromechanical vibrator. The selection of the edge-punched cards as the information carrier is a compromise solution which permits in some degree satisfying both of the requirements formulated above – following the current publications and retrospective search. Code information (notches on the card edge) will serve for the retrospective search. Information presented in ordinary language will be used for the current monitoring. The bibliographic description of the publication and a summary or brief abstract are printed on the card. A microfilm with a complete text of the article may also be mounted in the card [3].

Filling out the card. Speed – this is one of the basic requirements imposed on the information system. The completed punched cards (with mounted microphotocopy) must arrive at the personnel interested in them within a few days after the appearance of the journal. In order to shorten the time expended in preparation of the card, the information (title, abstract, or summary) must be reprinted on the card from the journal word-by-word without translation into Russian. The following system for filling out the card is suggested: the typist reprints on the cards word-for-word the information and transmits them together with the journal to the abstracter – a highly qualified specialist on mathematical statistics. The abstracter scans the articles, codes their contents, and, if necessary, writes a very short summary – one or two lines. In contrast with abstracting, the coding takes very little time – the abstracter should not study carefully the contents of the article, it is sufficient to scan the article and gather what is being

talked about. Our operating experience shows that from one to two working days are required for coding and a brief summary of a journal of average size.

Key words and the coding system. Our suggested coding system is presented in the second part of this article. We code: 1) bibliographic information (year of publication, type of publication, language, first letter of surname of the first author); 2) the mathematical concepts (key words) which characterize the contents of the article; 3) field of application (for works of an experimental nature). The contents of the publication may be characterized by several mathematical concents, therefore they are coded using a direct code. The fields of application are coded using a combined code – here it is assumed that the publications, as a rule, will relate only to some one applied discipline.

Only 140 mathematical concepts may be coded in our system. Experience shows [4] that the ratio of the number of publications to the number of key words should not be greater than two orders. It appears that our information system will reach its reasonable upper limit after 5-7 years. Before that time it will be necessary to prepare for creating a descriptor information system with a large number of key concepts (of the order of 1000). As the information carrier in such systems use may be made of the so-called superposition punched cards [5]. We hope that specialized information-logic computers will be ready for production in the course of the next 5-10 years.

It is interesting to compare our system of key mathematical concepts (see the second part of this article) with the system of the International Journal of Statistical Abstracts (see Appendix 1). Our system was developed independently. In spite of this, there is no great difference between the systems. The classification of the International Journal is compiled from the point of view of the researchers interest-

ed primarily in the development of mathematical statistics as an independent scientific discipline. The key concept system of the American journal Quality Control and Applied Statistics (see Appendix 2) differs significantly both from our system and from the system of the International Journal of Statistical Abstracts. Here there is clearly the result of the unique formulation of the problem – the operation is considered primarily from the point of view of the statistician – the specialist on quality control. (Quality control is an engineering specialty which is widely used abroad. The American Society of Quality Control has 16,000 members and has its own periodic publications, etc.).

Comparative evaluation of the abstract journals. It is interesting to compare the information systems in existence at the present time from the point of view of the problem of interest to us.

A portion of the RZh Matematika is the section Probability Theory and Mathematical Statistics, Theoretical Cybernetics, which is also issued in the form of a separate volume. This is a wide ranging journal. It includes the following sections.

I. *Probability theory and mathematical statistics.*

1. Probability theory.

2. Mathematical statistics.

3. Application of probability-theoretic and statistical methods.

II. *Theoretical cybernetics.*

1. General questions.

2. Control system theory and its application.

3. Automatic control theory.

4. Information theory.

5. Operations research and mathematical economics.

6. Theory of mathematical machines.

7. Mathematical and cybernetic questions of biology and psycholo-

gy.

8. Mathematical questions of semiotics.

III. *Author index.*

In sections I.2 and I.3 there are located primarily abstracts on mathematical works of a methodological nature. The methodological works of a nonmathematical nature are not covered completely. The selection of such works is somewhat random. The abstracts are accompanied by code indices using the UDK (Universal Decimal Classification).

The International Journal of Statistical Abstracts* is a narrowly specialized journal, publishing mathematical studies of a methodological nature. The abstracts are arranged in the journals in accordance with the adopted system of key concepts. Paper of different color is used for the different sections. The abstracts are short (400 words is recommended by "UNESCO"). They are printed only on one side of the page. Each abstract is accompanied by code notation which characterizes the contents of the article. The journal may be easily unbound and rebound by basic sections. We mentioned previously that 417 journals were examined for the selection of the publications (in 1963). These journals are distributed by countries as follows (%): USSR-1.44; USA-29.98; Great Britain-11.99; India-8.87; in German-7.43; Italy-7.19; France-3.60; Australia-3.12; Japan-3.12; other: Netherlands-2.64; Hungary-1.68; Egypt-1.22; Czechoslovak SSR-0.72, etc.

The American abstract journal Quality Control and Applied Statistics is also of a narrowly specialized nature. As a rule, it publishes long abstracts – two and even sometimes three pages in length. It is assumed that the readers using such extensive abstracts will be able to obtain a clear idea of the paper without recourse to reading the original publication. The abstracts are presented in a strictly standardized form, having the following sections: heading, author, journal,

purpose of the study, summary, results.

Each abstract is accompanied by code notations which characterize the contents of the article. (A characteristic of the journal is speed. As a rule, the abstracts appear with a delay of less than six months.)

Comparing these three journals, we can draw the following conclusions.

1. In the RZh Matematika the abstracts on works of a theoretical nature are written, as a rule, on a high level. This cannot be said of certain abstracts written on the statistical studies of a methodological nature. As a rule, in the foreign journals mentioned above the mathematical studies of a methodological nature are abstracted on a higher level. Mathematical statistics has today become a broadly branched discipline. In many branches of this science we do not have adequately qualified abstractors.

2. We must recognize the very successful, clear arrangement of material by headings used in the International Journal of Statistical Abstracts. It is considerably easier to follow the development of individual branches of mathematical statistics in the International Journal than in the RZh Matematika.

3. The specially developed titling systems of the foreign journals are considerably more convenient than that used in the UDK system which we have adopted. We get the impression that the UDK indices in application to works on mathematical statistics may be considered as noisy components of the abstracts rather than anything else.

4. In 1963 the following number of abstracts were published in the three journals considered.

RZh Matematika. Individual issue Probability Theory and Mathematical Statistics, Theoretical Cybernetics. Sections I.2 and I.3	1273

International Journal of Statistical Abstracts, Methodological Series	1000
Quality Control and Applied Statistics	547

Fewer abstracts are published in the foreign abstract journals than in the corresponding sections of the RZh Matematika. But this does not mean that the publications of the foreign journals may be considered as subsets of the publications of RZh Matematika. In these journals, particularly in the journal Quality Control and Applied Statistics, there are often included abstracts of articles from publications which VINITI does not receive. In the near future we plan to make a special study of the fields where the three abstract journals do not overlap.

Analysis of growth curves. Using the information system, it will be possible to follow the development of individual trends of the mathematical theory of experimentation, make a comparative estimate of the development of mathematical statistics in the Soviet Union, make certain predictions, and present individual recommendations.

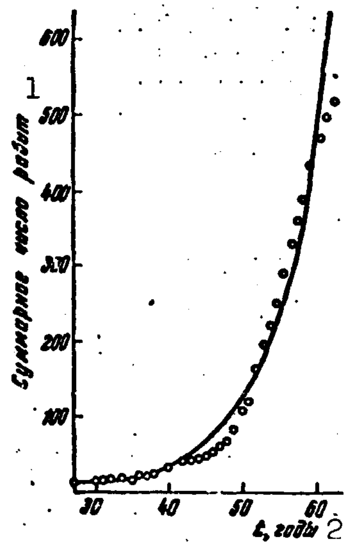
The growth curves constructed from the individual sections of the mathematical theory of experimentation may be of considerable interest. One such curve is shown in the figure. Here the years are plotted along the abscissa axis, the ordinate axis is the total number of methodological works relating to the planning of screening experiments in biology. This trend of mathematical statistics is developing in connection with the problems of crop husbandry, animal husbandry, investigation of pharmaceutical preparations, herbicides, etc. The curve was plotted from the bibliography of [6]. The curve is approximated quite well by an exponential. A sharp deviation from the exponential growth is seen in the war years, which is quite natural. The last points also deviate somewhat from the curve. Two hypotheses may be suggested to explain

this phenomenon: the rate of growth is diminishing – the exponential is transitioning into a logistic curve; the bibliography has not been accumulated sufficiently completely in recent years. To be conservative, it would be wiser to accept the second hypothesis in the present case. Here the relative rate of growth is 12.95%. This corresponds to doubling the total number of works in 5.4 years. A similar growth curve was constructed for the methodological works on factor experiment (in the sense in which this is understood in psychological studies). Here we used the extensive bibliography presented in [7]. In this case the relative rate of growth was 8.59%, which corresponds to doubling the number of works in 8 years. Finally, using the materials of our bibliography we plotted a growth curve for the total number of works (methodological and experimental) on the planning of extremal experiments. Here the relative growth rate was 43.6%, which corresponds to doubling in 1.6 years. It is pertinent to mention that the growth rate of the number of publications in the major divisions of science such as physics, chemistry, biology varies from 5 to 7%, which corresponds to doubling the number of works in 10-15 years [8]. The growth rate of works on planning of extremal experiments must be considered to be very high. It is interesting to compare the growth curves for two trends in statistics – planning of extremal and factor experiments. In the first case the total number of works of a methodological nature is an order less, and the growth rate is higher by nearly a factor of 5. This is explained by the fact that the first of these trends began to develop in 1913, the second started only in 1951. Scanning the journals, investigators usually judge the development of new trends by the total number of publications, and not by the rate of growth. It is clear that in this case many rapidly developing trends may remain unnoticed.

After the new trends in the mathematical theory of experimentation

have been identified, it is natural to turn to the second step - compile surveys and give a vigorously justified evaluation of the new trends. In certain cases monographs will be written as the next step following the reviews.

Using the information system, we will also be able to obtain useful information on the preparation and retraining of personnel, on the activity of the scientific societies, on the organization of scientific research and consultative centers, on seminars, conferences, etc.



Growth curve of number of works on screening experiment. 1) Total number of works; 2) t, yrs.

II. KEY WORDS AND THE CODE SYSTEM IN THE PUNCHED CARD FILE "MATHEMATICAL THEORY OF EXPERIMENTATION"

Yu.P. Adler, I.F. Aleksandrova, G.G. Vorob'yev, Yu.V. Granovskiy, Ye.V. Markova, and V.V. Nalimov

General principles. 1. The punchcard file uses the "manual" search methods which are suitable for 200-1000 searches per month with a number of punchcards for a single search not exceeding 10,000-40,000. Considering the growth rate of the number of works in this field, we can expect that the card file will not exceed these limitations in the course of the next 10 years.

2. The coding of the bibliographic and subject information is accomplished by cutting openings (edge notching or connecting openings with one another) corresponding to the given characteristic.

3. The punchcards used are similar to those of the RZh Metallurgiya, with minimal changes in the format which are noted below. All the publications are classified under 11 subject headings, each of which has 10 subheadings.

4. The internal field of the card is divided into two equal parts. The left part and the corresponding part on the reverse side are intended for the abstract text. The right side remains free. The production of transparent photographic microcards (up to 40 pages), which will be mounted in the punchcard, will be organized in the future.

Coding. 1. The search labels (key words) are divided into two categories: parallel (for example, subject) and mutually exclusive (for example, year of publication). Depending on this division we use respectively direct and combination codes. Each label is coded by a number. The direct code places in correspondence with the label (number) a single notch, while the combination code uses a combination of several notches. In a given card file the direct code is used for coding the mathematical concepts. The combination code is used for coding certain bibliographic information and information on fields of application.

2. The punchcard adopted permits three versions of standard notches:

1) a shallow notch – an opening of the outer row is brought out to the edge;

2) a deep notch – an opening of the inner row is brought out to the edge;

3) a combining notch (slit) connects a pair of openings; one inner and one outer.

3. For the direct code the shallow notches correspond to the labels which are coded by the openings of the outer row, the connecting openings (slits) correspond to the labels which are coded by the openings of the inner row, and the deep notches correspond simultaneously to the outer and inner openings.

4. When searching for a subheading of the outer row, the needle is inserted in the outer opening. In this case both the cards having only this label and the cards containing simultaneously an opening of the inner row fall out.

When searching for a subheading of the inner row the needle is inserted first in the opening of the inner row. After separating the cards with a deep notch the remaining stack of cards is grasped by the left hand and the needle which is removed is now inserted in any opening of the inner row and the separated cards are selected. In other cases the search is conducted similarly.

5. Only two types of notches are used for the combination code: shallow and deep. For all the labels on the upper edge of the card (other than the first letter of the author surname, see below) we use the key 7-4-2-1, which permits establishing the correspondence between the number (code) and a pair of notches.

When it is necessary to code 100 labels rather than 10, we can use two fields with the code 7-4-2-1. In this case ones are coded in one field, and tens in the other.

Filling out the cards. First all the data are entered on a model of the card which is printed on cheap newsprint. The code system is noted on the model. The entry of data on the model is accomplished directly by research personnel. The data are transferred from the model to the punchcard, which does not contain the expanded code system — the card area is retained for mounting a microfilm of the publication.

Transfer of the data from the model to the card and duplication of the cards may be performed by technicians.

The system "model with expanded code - card without code" is convenient in operation. By using this technique we can significantly simplify the most tedious process - the coding, without encumbering the basic information carrier, the card, with the code notations.

Along with the coded information, the following data are entered in the usual fashion on the card:

1) publication data of the article (translation of the title of the article is not mandatory);

2) indication of the abstract journal if a search is being made of the journal;

3) indication on the availability of a photocopy of the journal article at the information center;

4) indication of the location of a book and its library number.

An abstract or a short summary is also printed on the card.

Use of the card file. 1. It is advisable that the cards be stored in stacks of 1000 units in drawers with dimensions which fit the cards precisely. The size of a working stack is 200-250 cards.

2. When processing the working stack it is oriented with the operating edge up and aligned. Then, grasping the stack with the left hand near the required opening, the needle is introduced with the right hand (palm up). Then the left hand is shifted to the left edge of the stack, this edge is brought toward the operator and then moved away. This spreads out the cards on the needle and permits the free cards to fall out, which are gathered by the left hand. The piles of cards which have fallen free and those which remain must be carefully verified to be sure that there are no cards randomly stuck together and which have not fallen out and also to verify that there are no cards which have

fallen out because of edge tears.

3. This card file is a single unit which is assigned the code 1. With further development of the card file we plan to organize several units on related questions. For storage each unit may be arranged with respect to any subject tag or in an arbitrary order.

4. The search plan must be arranged so that in the first selection we obtain a minimal number of cards. With a reduction of the number of cards which fall out to 20-30 it is advisable to change over to the conventional manual sorting.

Adding to and correcting the system. 1. The search system provides the possibility of supplementing by using the vacant combinations in the code, free openings in the fields, and free fields if there are any. The three corner openings should not be used, since they enable the extraction from the card file of accidentally inverted cards.

2. When correcting the system (for example, when improving the classification) the subject headings are partially or completely assigned new values. For combined storage of old and new cards a signal notch is made in one of the unused openings of the new cards. For a slight change of the model the cards with obsolete notches are removed from the card file and the unnecessary notches in these cards may be accurately patched or new notches may be cut.

Equipment. 1. The necessary equipment includes:

1) drawers which are precisely fitted to the card width and are designed for 1000 cards;

2) a tumbler-spitsa with one pointed end and a handle on the other end. The material is stainless steel. Dimensions are: length of working section (without handle) 200-300 mm, diameter of working section 1.5-2 mm;

3) a manual punch for the notches (of the railroad ticket punch

type). This may be replaced by scissors and a safety razor blade.

2. More complex equipment:

1) mechanical punch for simultaneous punching of one opening in a stack of cards;

2) a coding machine for simultaneously punching all the necessary openings;

3) needles with a catch (of the safety-pin type);

4) a clip for manual sorting using several openings at once.

Other equipment may also be used.

3. The Microfot reader with a special holder is used for reading the aperture punchcards (with microcopies of the text inserted).

Code system. 1. The index for the first letter of the author surname is similar to the index of the RZh Metallurgiya [9].

2. Index for the publication type. The search with respect to type of publication may be performed in two aspects: with respect to form (article, book, etc.) and with respect to content (theoretical, applied, etc.).

The code for the publication type with respect to content is shown in Table 1.

The code for the publication type with respect to form is shown in Table 2. The key 4-2-1 is used (part of the key 7-4-2-1). It permits establishing the correspondence between a pair of openings and the numbers from 1 to 6.

3. Language index. For the language index we use the card field which in the standard cards of the RZh Metallurgiya is denoted by the term "branch." The code for the language index is shown in Table 3.

4. The chronological index within the limits of one decade is similar to the index of the RZh Metallurgiya of VINITI of the Academy of Sciences USSR (Table 4).

TABLE 1

Publication Type by Content

1 Признак	2 Код	Тип выреза 3
4 Теоретическая математическая	1	5 Мелкий
6 Теоретическая физико-химическая	2	7 Щель
8 Прикладная	3	5 Мелкий
9 Обзор, библиография	4	7 Щель

Remarks. A publication of methodological nature is coded by two shallow notches: 1 + 3. If the publication is not a survey paper, but contains a long bibliography, it is in addition coded with a notch 4 (slit).

1) Label; 2) code; 3) notch type; 4) theoretical mathematics; 5) shallow; 6) theoretical physico-chemical; 7) slit; 8) applied; 9) review, bibliography.

TABLE 2

Publication Type by Form

1 Признак	2 Код	Сочетание отверстий 3	Тип выреза 4
5 Статья	1	—	Глубокий 6
7 Книга, монография, труды конференции	2	—	•
8 Диссертация	3	1+2	Мелкий 9
10 Отчет	4	—	Глубокий 6
11 Словарь, пособие по переводу	5	1+4	Мелкий 9
12 Справочник	6	2+4	Глубокий 6

1) Label; 2) code; 3) combination of openings; 4) notch type; 5) articles; 6) deep; 7) book, monograph, conference transactions; 8) dissertation; 9) shallow; 10) report; 11) dictionary, translation aid; 12) handbook.

TABLE 3

Language Index

Язык 1	Код 2	Сочетание отверстий 3	Тип выреза 4
5 Английский	1	—	Глубокий 6
7 Русский	2	—	»
8 Немецкий	3	1+2	Мелкий 9
10 Французский	4	—	Глубокий 6
11 Японский	5	1+4 12	Мелкий 9
	6	2+4 (резерв)	»
	7	— (резерв)	»
	8	1+7 (резерв)	»
	9	2+7 (резерв)	»
13 Прочие	0	4+7	»

1) Language; 2) code; 3) combination of holes; 4) notch type; 5) English; 6) deep; 7) Russian; 8) German; 9) shallow; 10) French; 11) Japanese; 12) reserve; 13) other.

TABLE 4

Chronological Index

Год 1	Код 2	Сочетание отверстий 3	Тип выреза 4	Год 1	Код 2	Сочетание отверстий 3	Тип выреза 4
1961	1	—	5 Глубокий	1966	6	2+4	Мелкий 6
1962	2	—	» 6	1967	7	—	Глубокий 5
1963	3	1+2	Мелкий 5	1968	8	1+7	Мелкий 6
1964	4	—	Глубокий	1969	9	2+7	»
1965	5	1+4	Мелкий 6	1970	0	4+7	»

1) Year; 2) code; 3) combination of holes; 4) notch type; 5) deep; 6) shallow.

For coding publications from 1951 to 1960 we use the same system of codes, but in this case in addition the closest upper hole to the field on the right is punched. The publications prior to 1951 are coded by a slit alongside the field.

5. The subject index is shown in Table 5. It consists of 11 major headings 1.0, 1.1, ..., 2.0, each of which is divided into subheadings which are certain concepts which are elements (words) of the system information language.

On the card there remains room for two decimal headings 2.1, 2.2. These headings remain in reserve – they will be filled in as the index-

ing system is expanded.

Fields of application. In the great majority of cases each of the publications may relate only to some general field of application. Therefore for this division we use the combination code which has particularly high capacity: 100 concepts may be arranged on 60 holes. For the section "fields of application" we use that part of the field which in the standard cards of the RZh Metallurgiya has the label "country." The right side of this field is used for the numerical code 0-9, the left side is used for denoting the first digit of a binary number. The code is shown in Table 6.

Examples. Number 26 denotes "Study of particles, grains, etc." The digital and literal codes indicate that in the tens field there will be punched two shallow notches 2 + 1, and in the ones field there will also be two shallow notches 2 + 4. The number 30 denotes "Medicine." In this case in the tens field there is punched a single deep notch 4, while in the ones field there are two shallow notches 4 + 7.

TABLE 5

Subject Index

	Подрубрика 1	Код 2	Тип выреза* 3
4	**1.0. Общие вопросы**		
5	История, прогнозы, персоналии	0	М
6	Некоторые вопросы философии и методологии	1	Щ
7	Организация научных исследований, производства и администрирования. Подготовка кадров	2	М
8	Вопросы терминологии и стандартизации	3	Щ
9	Документалистика	4	М
10	Представление результатов эксперимента	5	Щ
11	Конференции и совещания	6	М
12	Таблицы (кроме таблиц случайных чисел)	7	Щ
13	Резерв	8	М
14	Прочие	9	Щ
15	**1.1. Программирование, вычислительная математика**		
16	Общие вопросы	0	М
17	Ручные средства (линейка, номограммы, мех. и электр. маш.)	1	Щ
18	Электронные вычислительные машины	2	М
19	Применение технических средств	3	Щ
20	Программирование	4	М
21	Автоматизация программирования. Машинные языки	5	Щ
22	Общие вопросы вычислительной математики	6	М
23	Вычислительные методы линейной алгебры	7	Щ
24	Другие вычислительные методы	8	М
25	Моделирование технологических и измерительных процессов на ЭЦВМ	9	Щ
26	**1.2. Техника эксперимента**		
16	Общие вопросы	0	М
27	Выбор параметра оптимизации	1	Щ
28	Активный эксперимент	2	М
29	Пассивный эксперимент	3	Щ
30	Пилотные установки	4	М

*M is a shallow notch, Shch is a slit.

TABLE 5 (continued)

	Подрубрика 1	Код 2	Тип выреза 3
31	Сбор информации о работающем процессе	5	Щ
32	Формализация априорной информации. Ранжирование	6	М
33	Метрологические вопросы	7	Щ
34	Воспроизводимость, ошибки прямых и косвенных измерений	8	М
13	Резерв	9	Щ
35	**1.3. Объекты исследования**		
15	Общие вопросы	0	М
36	Лабораторные объекты	1	Щ
37	Производственные процессы	2	М
38	Автоматизированные объекты	3	Щ
39	Экономические системы	4	М
40	Проектирование	5	Щ
41	Большие системы (включая ПЕРТ)	6	М
42	Автоматизация измерений	7	Щ
13	Резерв	8	М
14	Прочие	9	Щ
43	**1.4. Методы исследования I**		
16	Общие вопросы	0	М
44	Детерминистические методы	1	Щ
45	Стохастические методы	2	М
46	Смешанные методы	3	Щ
47	Исследование операций	4	М
48	Теория надежности	5	Щ
49	Линейное программирование	6	М
50	Нелинейное программирование	7	Щ
51	Динамическое программирование, вариационный анализ, метод Понтрягина	8	М
52	Теория игр. Принятие решений	9	Щ
53	**1.5. Методы исследования II (вероятностные аспекты)**		
16	Общие вопросы	0	М
54	Функции распределения	1	Щ
55	Нормальное распределение	2	М

TABLE 5 (continued)

Подрубрика 1	Код 2	Тип выреза 3
5 6 Логнормальное распределение	3	Щ
5 7 Прочие распределения	4	М
5 8 Мощность критерия	5	Щ
5 9 Свойства оценок	6	М
6 0 Критерии (χ^2, t, r, F и т. п.)	7	Щ
6 1 Преобразование переменных	8	М
6 2 Метод наименьших квадратов	9	Щ
6 3 **1.6. Методы исследования III (статистические методы)**		
1 6 Общие вопросы	0	М
6 4 Дисперсионный анализ	1	Щ
6 5 Регрессионный анализ	2	М
6 6 Корреляционный анализ	3	Щ
6 7 Ранговая корреляция	4	М
6 8 Непараметрическая статистика	5	Щ
6 9 Последовательный анализ (включая алгоритм Вальда)	6	М
7 0 Факторный анализ (в том числе как он понимается в психологии)	7	Щ
7 1 Дискриминантный анализ	8	М
7 2 Выделение аномальных наблюдений	9	Щ
7 3 **1.7. Методы планирования эксперимента (кроме дисперсионного анализа)**		
1 6 Общие вопросы	0	М
7 4 Факторные планы; дробные планы; полные и неполные блоки	1	Щ
7 5 Ротатабельные планы	2	М
7 6 Исследование поверхности отклика	3	Щ
7 7 Отсеивающий эксперимент (включая метод случайного баланса)	4	М
7 8 Эволюционное планирование. Адаптационная оптимизация	5	Щ
7 9 Планирование при изучении диаграмм состав — свойство	6	М
8 0 Планирование при уточнении значений параметров	7	Щ
8 1 Интерпретация результатов	8	М
8 2 Другие методы планирования	9	Щ

TABLE 5 (continued)

Подрубрика 1	Код 2	Тип выреза 3
8 3 **1.8. Методы теории случайных функций**		
1 6 Общие вопросы	0	М
8 4 Таблицы случайных чисел; генерирование случайных чисел; исследование случайных последовательностей; белый шум	1	Щ
8 5 Обобщенный гармонический анализ (автокорреляционная функция, взаимокорреляционная функция, спектральная плотность)	2	М
8 6 Определение динамических характеристик объекта	3	Щ
8 7 Фильтрация и сглаживание	4	М
8 8 Прогнозирование	5	Щ
8 9 Выделение сигнала	6	М
9 0 Граница обнаружения	7	Щ
9 1 Синтез оптимальных систем автоматического регулирования	8	М
9 2 Классификация процессов	9	Щ
9 3 **1.9. Контроль качества**		
1 6 Общие вопросы	0	М
9 4 Приемочный контроль	1	Щ
9 5 Время жизни	2	М
9 6 Планирование выборок	3	Щ
9 7 Межлабораторные испытания	4	М
9 8 Контрольные диаграммы	5	Щ
9 9 Рандомизация	6	М
ГОСТЫ и другие стандарты 1 0 0	7	Ш
1 3 Резерв	8	М
1 3 Резерв	9	Ш
1 0 1 **2.0. Смежные вопросы (эта рубрика попадает на поле «материалы»)**		
1 6 Общие вопросы	0	М
Теория информации 1 0 2	1	Щ
Теория массового обслуживания 1 0 3	2	М
Математическая логика 1 0 4	3	Щ
Многомерная геометрия и топология (в применении к планированию эксперимента)	4	М
1 0 5		

TABLE 5 (continued)

	1 Подрубрика	2 Код	3 Тип выреза *
106	Теория графов	5	Щ
107	Комбинаторный анализ	6	М
108	Бионика. Психологические шкалы. Психофизиологические методы переработки информации	7	Щ
109	Методы Монте-Карло	8	М
110	Дифференциальные и интегральные уравнения	9	Щ

1) Subheading; 2) code; 3) notch type*; 4) general questions; 5) history, predictions, personalities; 6) some questions of philosophy and methodology; 7) organization of scientific research, production, and administration. Personnel preparation; 8) questions of terminology and standardization; 9) documentalistics; 10) presentation of experiment results; 11) conferences and meetings; 12) tables (other than tables of random numbers); 13) reserve; 14) other; 15) 1.1. Programming, computational mathematics; 16) general questions; 17) manual equipment (slide rule, nomograms, mechanical and electrical computers); 18) electronic computers; 19) use of technical facilities; 20) programming; 21) automation of programming. Machine languages; 22) general questions of computational mathematics; 23) computational methods of linear algebra; 24) other computational methods; 25) simulation of technological and instrumentation processes on electronic digital computers; 26) 1.2. Experimental technique; 27) selection of optimizing parameter; 28) active experiment; 29) passive experiment; 30) pilot installations; 31) collection of information concerning operating process; 32) formalizing *a priori* information. Ranking; 33) metrological questions; 34) reproducibility, errors of direct and indirect measurements; 35) 1.3. Study objects; 36) laboratory objects; 37) production processes; 38) automated objects; 39) economic systems; 40) design; 41) large systems (including PERT); 42) measurement automation; 43) 1.4. Methods of investigation I; 44) deterministic method; 45) stochastic method; 46) mixed methods; 47) operations research; 48) reliability theory; 49) linear programming; 50) nonlinear programming; 51) dynamic programming, variational analysis, Pontryagin method; 52) game theory. Decision making; 53) 1.5. Methods of investigation II (probability aspects); 54) distribution functions; 55) normal distribution; 56) lognormal distribution; 57) other distributions; 58) criteria power; 59) properties of estimates; 60) criteria (χ^2, t, r, F, etc.); 61) transformation of variables; 62) method of least squares; 63) 1.6. Methods of investigation III (statistical methods); 64) dispersion analysis; 65) regression analysis; 66) correlation analysis; 67) rank correlation; 68) nonparametric statistics; 69) sequential analysis (including the Wald algorithm); 70) factor analysis (including how it is understood in psychology); 71) discriminant analysis; 72) identification of abnormal observations; 73) 1.7. Methods of experiment planning (other than dispersion analysis); 74) factor plans, fractional plans, complete and incomplete blocks; 75) rotatable plans; 76) study of response surface; 77) screening experiment (including the method of random balance); 78) evolu-

TABLE 5 (continued)

tional planning. Adaptive optimization; 79) planning in the study of composition-property diagrams; 80) planning for refining values of parameters; 81) interpretation of results; 82) other planning methods; 83) 1.8. Methods of random function theory; 84) Tables of random numbers, generation of random numbers, study of random sequences, white noise; 85) generalized harmonic analysis (auto-correlation function, cross-correlation function, spectral density); 86) determination of dynamic characteristics of object; 87) filtering and smoothing; 88) prediction; 89) signal identification; 90) detection boundary; 91) synthesis of optimal automatic control systems; 92) process classification; 93) 1.9. Quality control; 94) acceptance inspection; 95) lifetime; 96) sample planning; 97) inter-laboratory testing; 98) control diagram; 99) randomization; 100) GOSTY and other standards; 101) 2.0. Related questions (this heading falls in the "materials" field); 102) information theory; 103) queueing theory; 104) mathematical logic; 105) multi-dimensional geometry and topology (in application to experiment planning) 106) graph theory; 107) combinatorial analysis; 108) bionics. Psychological scales. Psychological and physiological methods of information processing; 109) Monte Carlo methods; 110) differential and integral equations.

TABLE 6

Fields of Application

a №№ п/п.	b Подрубрика	c Код	d Тип выреза *
1	Общее (применение в разных областях)	1/1	Г, Г
2	Аналитическая химия	2/1	Г, Г
3	Неорганическая химия	1+2/1	М, Г
4	Органическая химия	4/1	Г, Г
5	Химия высокомолекулярных соединений. Полимеры	1+4/1	М, Г
6	Физическая и коллоидная химия	2+4/1	М, Г
7	Химия нефти	7/1	Г, Г
8	Геохимия и геология	1+7/1	М, Г
9	Биохимия и биология	2+7/1	М, Г
10	Фармацевтическая промышленность	4+7/2	М, Г
11	Производство каучука и резины. Шинная промышленность	1/2	Г, Г
12	Лесная промышленность (вместе с целлюлозно-бумажной)	2/2	Г, Г
13	Пищевая промышленность	1+2/2	М, Г
14	Химия (прочие области)	4/2	Г, Г
15	Техническая физика	1+4/2	М, Г
16	Полупроводниковая металлургия	2+4/2	М, Г
17	Черная металлургия	7/2	М, Г
18	Цветная и редкая металлургия	1+7/2	М, Г
19	Стройматериалы	2+7/2	М, Г

(continued)

a №№ п/п.	b Подрубрика	c Код	d Тип выреза
20	Угольная промышленность	4+7/2+1	М, М
21	Электро- и светотехническая промышленность	1/2+1	Г, М
22	Электроника	2/2+1	Г, М
23	Машиностроение	1+2/2+1	М, М
24	Прочие отрасли промышленности	4/2+1	Г, М
25	Применение вычислительных устройств в промышленности	1+4/2+1	М, М
26	Исследование частиц, зерен и пр.	2+4/2+1	М, М
27	Юриспруденция	7/2+1	Г, М
28	Психология	1+7/2+1	М, М
29	Социология	2+7/2+1	М, М
30	Медицина	4+7/4	М, Г
31	Спектральный анализ (оптический и рентгеновский)	1/4	Г, Г
32	Гамма-спектроскопия	2/4	Г, Г
33	Фотографический процесс	1+2/4	М, Г
34	Сельское хозяйство, агробиология	4/4	Г, Г
35	Связь	1+4/4	М, Г
36	Астрономия	2+1/4+2	М, М
37	Испытание материалов	2+1/7	М, Г

* Г is a deep notch; М is a shallow notch.

a) No.; b) subheading; c) code; d) notch type*; 1) general (application in various fields); 2) analytic chemistry; 3) inorganic chemistry; 4) organic chemistry; 5) chemistry of high-molecular compounds. Polymers; 6) physical and colloid chemistry; 7) petroleum chemistry; 8) geochemistry and geology; 9) biochemistry and biology; 10) pharmaceutical industry; 11) production of caoutchouc and rubber. Tire industry; 12) wood-products industry (together with cellulose and paper industry); 13) foodstuffs industry; 14) chemistry (other fields); 15) engineering physics; 16) semiconductor metallurgy; 17) ferrous metallurgy; 18) nonferrous and rare-metal metallurgy; 19) construction materials; 20) coal industry; 21) electrical and illumination industry; 22) electronics; 23) machine construction; 24) other branches of industry; 25) application of computers in industry; 26) study of particles, grains, etc.; 27) jurisprudence; 28) psychology; 29) sociology; 30) medicine; 31) spectral analysis (optical and x-ray); 32) gamma spectroscopy; 33) photographic process; 34) agriculture, agrobiology; 35) communications; 36) astronomy; 37) material testing.

REFERENCES

1. J.W. Tukey, The Future of Data Analysis. The Annals of Mathematical Statistics, 1962, 33, No. 1, 1.

2. M.M. Kessler. Bibliographic Coupling Between Scientific Pzpers. American Documentation, 1963, 14, No. 1, 10.

3. S.A. Lipkin and G.G. Vorob'yev. Fotograficheskiye metody nakopleniya laboratornoy informatsii [Photographic Methods of Storing Laboratory Information]. Zavodskaya laboratoriya, 1962, 28, No. 10, 1221.

4. G.G. Vorob'yev. Problema dokumental'noy informatsii [The Problem of Documental Information]. This Collection.

5. Perforirovannyye karty i ikh primeneniye v nauke i tekhnike [Punchcards and their Application in Science and Engineering]. Mashgiz, 1963.

6. W.T. Federer. Procedures and Designs Useful for Screening Material in Selection and Allocation, with Bibliography. Biometrics, 1963, 19, No. 4, page 553.

7. H. Harry. Modern Factor Analysis. Univ. Chicago Press, 1960, 469 pp.

8. G.E. Bleduts, V.V. Nalimov, and N.I. Styazhkin. Nauchnaya i tekhnicheskaya informatsiya kak odna iz zadach kibernetiki [Scientific and Technical Information as One of the Problems of Cybernetics]. UFN, 1959, 69, No. 1, page 13.

9. Instruktsiya po vvedeniyu perfokartoteki RZh "Metallurgiya" [Instructions for Introduction of the Punched Cards of the Abstract Journal "Metallurgy"]. VINITI, 1963-1965.

Manuscript Page No.	[Footnotes]
144	International Journal of Abstracts. Statistical Theory and Methods. Olivers and Boyd. Great Britain.
145	Abstract Service — Quality Control and Applied Statistics. Executive Sciences Institute, Inc., Whippany, New Jersey, USA.
150	B.V. Gnedenko represents the Soviet Union in this journal.

[Transliterated Symbols]

143	РЖ = RZh = Referativnyy Zhurnal = Abstract Journal
145	ВИНИТИ АН = VINITI AN = Vsesoyuznyy institut nauchnoy i tekhnicheskoy informatsii Akademii Nauk = Academy of Sciences All-Union Institute of Scientific and Technical Information
150	УДК = UDK = Universal'naya desyatichnaya klassifikatsiya = Universal Decimal Classification
163	М = M = melkiy = shallow
163	Щ = Shch = shchel' = slit
164	ГОСТы = GOSTy = Gosudarstvennyye Obshchesoyuznyye Standarty = All-Union State Standards
167	Г = G = glubokiy = deep

Appendix 1

CLASSIFICATION OF STATISTICAL METHODS IN THE INTERNATIONAL JOURNAL OF STATISTICAL ABSTRACTS

0. Mathematical methods

0. General articles

1. Solution of equations

2. Methods of constructing curves from points

3. Interpolation and quadrature

4. Special functions and their transformations

5. Functional relations

6. Determinants and matrices

7. Game theory

8. Programming techniques

9. Reserve

1. Probability theory

0. General articles

1. Probability calculation

2. Parameters of a general ensemble (individual estimates)

3. Problems of combinatorics

4. Geometrical probability

5. Limit theorems

6. Convergence in probability

7. Stochastic approximation

8. Theory of decision making and preference function

9. Reserve

2. Distribution

0. General articles

1. Descriptive properties

2. Transformations of variables

3. Normal and lognormal

4. Binomial, multivariate, normal, and hypergeometric

5. Poisson, exponential, negative binomial, logarithmic, and the distribution which depends on random parameters

6. Rectangular and extremal value

7. Pearson and "time series"

8. Truncated and mixed

9. Other distributions

3. Sampling distributions

0. General articles

1. t, z, F, and χ^2 distributions

2. Noncentral distributions

3. "Studentization"; normal approximations

4. Quadratic forms

5. Correlation and regression coefficients

6. Ordering and statistical scales

7. Descriptive statistics

8. Order statistics

9. Multifactor problems

4. Estimates

0. General articles

1. Properties of estimates

2. Types of estimates: fundamental theory

3. Individual estimates: the point

4. Same: the interval

5. Inequalities; admissible limits and regions

6. Methods which do not use distribution functions

7. Sequential methods

8. Multifactor problems

9. Finite ensembles – preliminary investigations

5. Hypothesis testing

0. General articles

1. Criteria properties

2. Single hypotheses

3. The dual sampling problem

4. The k-sampling problem

5. Sharply differing values

6. Criteria which do not use distributions

7. Sequential criteria

8. Multifactor problems

9. Reserve

6. Equations

0. General articles

1. Regression – the linear hypothesis

2. Correlation

3. Factor methods

4. Other multifactor methods

5. Ranking methods

6. Systems of equations; structure

7. Nonlinear equations – the logistic curve

8. Equation transformation – starting with two possibilities

9. Correspondence and conjugacy

7. Dispersion analysis

0. General articles

1. Model with constant factors

2. Model with random factors

3. Mixed and other models

4. Nonorthogonal observations and missing values

5. Nonstandard conditions (violation of premises

6. Covariational analysis

7. Multiple comparisons; processes of multiple decision making

8. Data ranking

9. Sequential methods

8. Sample planning (sampling plans)

0. General articles

1. Random samples; stratified samples, multi-stage samples

2. Samples with unequal probabilities

3. Multiple samples; dual samples

4. Natural populations

5. Nonsampling problems

6. Censored samples, systematic samples, and group samples

7. Nature and number of objects; cost and effectiveness

8. Acceptance inspection

9. Process control

9. Experiment planning

0. General articles

1. Planning with partitioning into blocks, complete and incomplete

2. Factor plans

3. Response surfaces

4. Nature of objects; number of replicas; cost and effectiveness

5. Paired comparisons and comparison problems

6. Preference criteria

7. Repetition of experiments; other planning

8. The weighting problem

9. Sensitivity problems

10. Theory of random processes and analysis of time series

0. General articles

1. Properties of individual processes

2. Estimate problems

3. Hypothesis verification

4. Queues and weighting-line theory

5. Information theory

6. Spectral analysis

7. Auto- and serial correlation

8. Multivariate processes

9. Investigations in biology, populations

11. Miscellaneous

0. General statistical methodology

1. Statistical tables and diagrams

2. Probability paper

3. Nomography and graphical methods

4. Machine methods: manual and with the use of punchcards

5. Machine methods: electronic computers

6. Machine methods: other

7. Monte Carlo methods

8. Indexing

9. History, biography, and bibliography

Appendix 2

CLASSIFICATION IN THE AMERICAN ABSTRACT JOURNAL "QUALITY CONTROL AND APPLIED STATISTICS"

100: Statistical industrial control

- 110: Cards for quality control
  - 111: Control cards for factors
  - 112: Control cards for properties
- 120: Specifications, tolerances, and process capabilities
- 130: Distribution functions in process control
- 190: Miscellaneous

200: Sample types and sampling plans

- 210: Sample types
  - 211: Mass sampling
  - 212: Discrete unit sampling
  - 219: Other
- 220: Sampling plans
  - 221: Selecting the correct sampling plans
  - 222: Plans for control of properties (Attributes)
  - 223: Plans for control of factors (Variables)
- 230: Census and survey
- 290: Miscellaneous

300: Control of quality control

- 310: Formulating the quality control task
- 320: Education in quality control
- 330: Organization of quality control

331: Personnel

340: Quality control administrative engineering

341: Recording and reports on quality

342: Standards and methods

343: Incentive plans

349: Other administrative methods

350: Economics of quality

351: User-vendor relationship

353: Quality standards

359: Miscellaneous

390: Miscellaneous

400: Mathematical statistics and probability theory

410: Estimate theory and statistical conclusions

420: Properties of distribution functions

421: Normal distribution

422: Poisson distribution

423: Binomial distribution

424: Mixed distribution

425: Distribution of extremal values

429: Other distributions

430: Probability theory

490: Miscellaneous

500: Experiments and the study of relationships

510: Significance criteria and confidence intervals

520: Planning and analysis of experiments

521: Experiment organization

522: Experiment planning

522.1: Complete blocks (factor plans, Latin squares, randomized blocks, etc.)

522.2: Incomplete blocks (Oudin squares, etc.)

522.3: Sequential plans

522.4: Fractional plans

522.5: Response surfaces

522.9: Other

523: Special methods of data analysis

523.1: Data conversion

523.2: Missing values

523.3: Prominent values and data processing

523.4: Dispersion analysis of ranked data

523.5: Table of characteristic conjugacy

523.9: Other

529: Miscellaneous

530: Correlation

531: Pairwise correlation

532: Multiple correlation

533: Rank correlation

534: Covariational analysis

539: Other

540: Construction of empirical curves

541: Linear regression

542: Nonlinear regression

543: Multiple regression

544: Orthogonal polynomials

545: Time series

546: Other

550: Simplified analysis methods

551: Nonparametric tests

552: Graphical analysis of data

553: Other

590: Miscellaneous

600: Application to equation.*

610: Operations research

611: Data collection

612: Special methods and their application

612.1: Programming

612.11: Transport problems

612.12: The assignment problem

612.13: The sequence problem

612.14: The traveling salesman problem

612.15: The porfolio problem

612.2: Information theory

612.3: Game theory

612.4: Queueing theory

612.5: Symbolic logic

612.6: Dynamic programming

612.7: Other mathematical methods

612.8: Modeling processes

612.9: Other

613: System analysis

620: Engineering industrial method

622: Compilation of graphs and control (production, inventory, etc.)

623: Work measurement and wage planning

625: Equipment replacement

629: Other

630: Commercial methods

631: Measurement and analysis

632: Prediction

633: Finance policies

634: Sales control

639: Other

690: Miscellaneous

700: Measurement and control

710: Measurement of quality characteristics

711: Physical properties

711.1: Length, volume

711.2: Mass, density

711.3: Capillary, colloidal, and surface properties

711.4: Mechanical properties, including rheological, ultimate strength, pressure

711.5: Electrical properties, including resistance, dielectric and magnetic properties

711.6: Optical properties, including external appearance, transparency, color, absorption spectrophotometry

711.7: Thermal properties, including temperature, resistance, expansion, heat capacity

711.9: Miscellaneous

712: Dynamic properties

712.1: Velocity, acceleration

712.2: Material flux, its calculation

712.9: Miscellaneous

713: Structural properties

713.1: Molecular structures

713.2: Internal structures

713.9: Miscellaneous

714: Chemical properties

714.1: Chemical composition

714.2: Chemical reactions

714.9: Miscellaneous

715: Atomic and nuclear properties

716: Aging properties and wear

717: Subjective characteristics

720: Process control and automation

730: Computers and special devices for applied statistics

800: Reliability of complex ensembles

810: Control of the reliability functions

811: Organization

812: Training

813: Data collection and reporting

814: Reliability standards

815: Reliability specifications

816: Vendor-buyer relationship

819: Miscellaneous

820: Mathematical theory of reliability

821: Models of probability chains

822: Distribution functions

823: Sampling plans for reliability estimate

824: Experiment planning and reliability

829: Miscellaneous

830: Aspects of system reliability

831: Systems with objective reliability

832: Human engineering aspect

833: Selection of components

834: Gain with redundancy

835: Suitability of predictions

839: Miscellaneous

840: Reliability improvement

841: Materials

842: Components

843: Ensembles

844: Systems

848: Tests, means, and development

849: Miscellaneous

850: The field of reliability and operation

890: Miscellaneous

:A Industry and production

:Aa Mechanical

:Ab Electrical and electronic

:Ac Chemical, petroleum, and pharmaceutical

:Ad Mining, ore-processing, and casting

:Ae Construction work

:Af Construction timber, wood and paper

:Ag Plastics and rubber

:Ah Leather

:Ai Textile

:Aj Ceramics and glass

:Ak Graphics

:Al Foodstuffs and beverages

:Am Packing industry

:An Rockets and spacecraft

:Az Miscellaneous

:B Trade and incentives

:C Banking, finance, and insurance

:D Research and development

:Da
:   Same as for :A
:Dz

:E Military, naval, and government operations

:F Social sciences, including psychology and education

:G Business correspondence

:H Biometrics (agronomy, genetics, etc.)

:I Medicine and related questions

:J Utilities and transport

:K Earch sciences (geophysics, meteorology, oceanography, etc.)

:L General questions of administration

:M Service organization (garages, restaurants, etc.)

:N Econometrics

:Y General applications

:Ya Estimate of correctness and accuracy

:Yb Testing for lifetime

:Yc Subjective criteria

:Z Other

Manuscript Page No.	[Footnote]
176	Actually, for the publications appearing in June of 1960 the abstracts on operations research are included in QCAS only to the extent to which the original works were concerned directly with quality control or reliability. For complete coverage of the literature on operations research the subscribers are referred to the service of Executive Sciences Institute, OR/MS.

Appendix 3

CLASSIFICATION SCHEME FOR BIBLIOGRAPHY OF THE INTERNATIONAL STATISTICAL INSTITUTE

0. General
   - 0.1 New periodic publications
   - 0.2 Conference transactions, memorial editions
   - 0.3 Bibliography
   - 0.4 Dictionaries, glossaries
   - 0.5 History of statistics and its bibliography
   - 0.6 Teaching of statistics
   - 0.9 Miscellaneous

1. Statistical theory
   - 1.0 General
   - 1.1 Probability
   - 1.2 Distribution functions
   - 1.3 Sampling distributions
   - 1.4 Estimates
   - 1.5 Hypothesis testing
   - 1.6 Equations
   - 1.7 Dispersion analysis
   - 1.8 Sampling plans
   - 1.9 Experiment planning
   - 1.10 Theory of random processes and analysis of time series
   - 1.11 Miscellaneous, including statistical tables and nomograms

2. Applications of statistics
   - 2.0 General
   - 2.1 Economics
   - 2.2 Sociology and psychology
   - 2.3 Demography and the science of insurance
   - 2.4 Biology and medicine
   - 2.5 Physical sciences and technology
   - 2.6 Operations analysis and the science of control
   - 2.9 Miscellaneous

3. Systematization of data
   - 3.0 General
   - 3.1 Legislation and organization
   - 3.2 Problems of definition and classification
   - 3.3 Statistics of basic materials
   - 3.4 Practical sampling problems
   - 3.5 Process data

3.6 Data representation

3.9 Miscellaneous

4. Statistical data

4.0 General

4.1 Populations and public health

4.2 Social statistics

4.3 Statistics of the cultural life and criminal statistics

4.4 Agrarian statistics

4.5 Industrial statistics

4.6 Trade and transport statistics

4.7 Financial statistics

4.8 Government accounting

4.9 Miscellaneous

A. Experiment

1. Experiment planning

2. Experiment conduct

3. Analysis of results

B. Preliminary inspection

1. Sampling plans

2. Systemization of data

3. Publication of data

4. Analysis of results